

GenCore version 5.1.1.3
Copyright (C) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 8, 2002, 16:42:31 ; Search time 6148.03 Seconds
(without alignments)
16383.279 Million cell updates/sec

Title: US-09-642-034-1
Perfect score: 3461
Sequence: 1 ctcgtgccgaattcgccacg.....aaattccttatcatcagcttg 3461

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues 4109280
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match %	Length	DB ID	Description
1	3461	100.0	3461	6	AX465588 Sequence
2	3391	98.0	3461	6	AX207205 Sequence
3	2761.2	79.8	2776	6	AX207207 Sequence
4	2716	78.5	2744	9	HSU41060 Homo sapien
5	2249	65.0	2404	6	176892 Sequence 3
6	2134.2	61.7	2236	9	BC008317 Homo sapi
7	1753.2	50.7	3287	10	AB071697 Mus muscu
8	1486.8	43.0	1597	6	AX017261 Sequence 1
9	1288	37.2	1310	6	176891 Sequence 1
10	1137.8	32.9	221941	9	AC091060 Homo sapi
11	1137.8	32.9	224788	2	AF001905 Homo sapi
12	804	23.2	160170	2	AF001158 Homo sapi
13	619.6	17.9	899	6	AX465590 Sequence
14	549.4	15.9	551	6	AX207216 Sequence
15	429.2	12.4	193437	2	AC126667 Rattus no
16	390.8	11.3	395	6	AX156237 Sequence
17	336.8	9.7	445	11	G24009 human STS W
18	319.2	9.2	337	6	AX393861 Sequence
19	319	9.2	5231	9	AB033091 Homo sapi
20	319	9.2	5322	6	AX405756 Sequence
21	282.8	8.2	61103	2	AC131276 Homo sapi
22	253.8	7.3	391	6	AX207209 Sequence
23	240	6.9	241	6	AX303040 Sequence
24	191.2	5.5	193437	2	AC126667 Rattus no
25	165	4.8	177	6	AX106381 Sequence
26	165	4.8	177	6	AX140672 Sequence
27	165	4.8	177	6	AX200532 Sequence
28	165	4.8	177	6	AX267188 Sequence
29	144	4.2	52684	2	AC014796 Drosophil
30	144	4.2	174712	3	AC010041 Drosophil
31	144	4.2	195148	3	AC010019 Drosophil
32	144	4.2	282267	3	AE003555 Drosophil
33	129.2	3.7	2174	10	BC021530 Mus muscu
34	128.4	3.7	2015	9	AB051127 Macaca fa
35	127.6	3.7	4573	6	AX409002 Sequence
36	127.6	3.7	4573	9	AX409002 Sequence
37	126	3.6	4610	6	AX078294 Sequence
38	126	3.6	4698	6	AX281784 Sequence
39	123.4	3.6	1383	6	AX118897 Sequence
40	123.4	3.6	1803	9	AF193052 Homo sapi
41	123.4	3.6	2604	9	AK027652 Homo sapi
42	123.4	3.6	3169	9	BC012125 Homo sapi
43	123.4	3.6	3246	9	AB040120 Homo sapi
44	122.6	3.5	1895	9	BC027884 Homo sapi
45	121.4	3.5	55577	2	AC014109 Drosophil

ALIGNMENTS

RESULT 1	AX465588	AX465588	3461 bp	DNA	linear	PAT 16-JUL-2002
LOCUS	Sequence 1	Sequence 1	from Patent WO0216939.			
DEFINITION	AX465588					
ACCESSION	AX465588.1	GI:21899891				
VERSION						
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE						
AUTHORS	Mack, D., Gish, K.C. and Wilson, K.E.					
TITLE	Methods of diagnosis of cancer and screening for cancer modulators					
JOURNAL	Patent: WO 0216939-A 1 28-FEB-2002;					

Pred. No. is the number of results predicted by chance to have a

EOS Biotechnology, Inc. (US)
 Location/Qualifiers
 1..3461
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="prostate/bladder/ovarian/bladder cancer protein
 human BCR4 CDNA"
 136..2405
 /note="human BCR4"
 /codon_start=1
 /protein_id="CAD42374.1"
 /db_xref="GI:21899892"
 /translation="MARLSSVILTLFTALSVTNPLHELKAAAFPOTTKEKISPNWESGI
 NVDLATSTQHYLQQLFQYRIENSSLSVEGFRKLQNIQIDKIRIHHDHDSHRS
 EHSDDHSHRSHSDHSDHSHHHAASGKNKRAKCPDHSDSDSGDKPKDV
 SOGGAHRPHSHASGRNNKVSASVSTVYNTVSGTHFLETETPRPKLLPKPKV
 SSSPPSVSTKSLRAGKRIKSNESVSEPRKSHKARIPKTYSLQIAHWGGGTAISI
 GIOPLNATFENYLCPAIIQNDARSLIHTESEKKAIFLHLLPHSHASHHSHSHE
 ISFLSLGLVILVLMNRVFEKLLSFLVALAVLGTLSGDAFLHLLPHSHASHHSHS
 EFAMKMRGFLSHLSSNISESAYFSDTWKGLTALGQDLFMLEVRIETLTKQFKDK
 KKKKKKRPENDDDVEIKKOLSKYESQLSTNEEKYDTRDTGLRADSPVSHFDSQ
 PALLEEVEVMAHAHQEVYNEVPRGCKNKCHSHFDHDTLQSDDLHIIHHHHYHLLS
 HHHNNHHPHSHRSISREBELDAGVATLAWMIIMDGLHNFSDGLAATGATFEGLS
 GLSTSVAVFCHLPELHGDVAVLLKAGMTVKOAVLYNALSAMLAYLGMATGATFIGHA
 ENVSMIFALTGLFMFVAVLVAVPEKMLNDASDHCCSRWGYFFLQAGMLLGGFLGI
 LISIFEHKIVFRNF"

BASE COUNT 1042 a 686 c 728 g 1005 t

Query Match	Score 3461;	DB 6;	Length 3461;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 3461;	Conservative	0; Mismatches	0; Indels
			0; Gaps
Qy 1	CTCTGCCGAATTCGGCAGCAGACCGCGTGTTCGCGCCTGGTAGAGATTTCCTGAAGACA	60	
Db 1	CTCTGCCGAATTCGGCAGCAGACCGCGTGTTCGCGCCTGGTAGAGATTTCCTGAAGACA	60	
Qy 61	CCAGTGGCGCGTGTGGAAACCAACCTCGCGCGTGGCGCGCGTGGGACAACGAGGCC	120	
Qy 61	CCAGTGGCGCGTGTGGAAACCAACCTCGCGCGTGGCGCGCGTGGGACAACGAGGCC	120	
Qy 121	CGCGAGACGAAGCGGCNATGGCGGAGGAGTTATCTGTAATCTTGATCTCTGACCTTGCCC	180	
Qy 121	CGCGAGACGAAGCGGCNATGGCGGAGGAGTTATCTGTAATCTTGATCTCTGACCTTGCCC	180	
Qy 181	TCTCTGTACAAATCCCTTTCATGAATCTGTAATCTTGATCTCTGACCTTGACCTGAGA	240	
Qy 181	TCTCTGTACAAATCCCTTTCATGAATCTGTAATCTTGATCTCTGACCTTGACCTGAGA	240	
Qy 241	AAATTAGTCCGAATGGGAATCTGGCAATTAATGTTGACCTGGCAATTTCCACACGCAAT	300	
Qy 241	AAATTAGTCCGAATGGGAATCTGGCAATTAATGTTGACCTGGCAATTTCCACACGCAAT	300	
Qy 301	ATCATCTACACAGCTTTCTACCGCTATCGGAAATAAATCTTTGTGACCTTGAGGGT	360	
Qy 301	ATCATCTACACAGCTTTCTACCGCTATCGGAAATAAATCTTTGTGACCTTGAGGGT	360	
Qy 361	TCAGAAATTTACTTCAAAATATAGCATAGATTAAGATTAAAGAAATCCATATACACCATG	420	
Qy 361	TCAGAAATTTACTTCAAAATATAGCATAGATTAAGATTAAAGAAATCCATATACACCATG	420	
Qy 421	ACCACGACCATGACTCAGACACGAGCATCACTCAGACCATGAGCGTCACTCAGACCATG	480	
Qy 421	ACCACGACCATGACTCAGACACGAGCATCACTCAGACCATGAGCGTCACTCAGACCATG	480	
Qy 481	AGCATCACTCAGACACGAGCATCACTTGACCATGATCATCTCTCCACCAATATCATG	540	
Qy 481	AGCATCACTCAGACACGAGCATCACTTGACCATGATCATCTCTCCACCAATATCATG	540	
Qy 541	CTGCTTCTGGTAAAATTAAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTGAG	600	
Qy 541	CTGCTTCTGGTAAAATTAAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTGAG	600	

Query Match	98.0%;	Score 3391;	DB 6,	Length
Best Local Similarity	99.4%;	pred. No. 0;		
Matches 3441; Conservative	0;	Mismatches	0;	Indels
			20;	Gaps

QY	1	CTCGTGGCGAATTCGGACAGAGACCGCGTGTTCGCGCTGGTAGAGATTTCGGAAGACA	60
DB	1	CTCGTGGCGAATTCGGACAGAGACCGCGTGTTCGCGCTGGTAGAGATTTCGGAAGACA	60
QY	61	CCAGTGGCGCGTGTGGAACCAAACTCCGCGCGTGGCGCGCGTGGGACAAACGAGGCC	120
DB	61	CCAGTGGCGCGTGTGGAACCAAACTCCGCGCGTGGCGCGCGTGGGACAAACGAGGCC	120
QY	121	GGGAGAGCAGAGCGCAATGGCGAGGAAGTTATCTGTAATCTTGATCCTGACCTTTGCC	180
DB	121	GGGAGAGCAGAGCGCAATGGCGAGGAAGTTATCTGTAATCTTGATCCTGACCTTTGCC	180
QY	181	TCTCTGTACAAAATCCCTTCATGAACTAAAGCAGCTGCTTTCCGCCACACACTGAGA	240
DB	181	TCTCTGTACAAAATCCCTTCATGAACTAAAGCAGCTGCTTTCCGCCACACACTGAGA	240
QY	241	AAATAGTCCGAATGGGAATCTGGCATTAATGTTGACTTGGCAATTTCCACACGGCAAT	300
DB	241	AAATAGTCCGAATGGGAATCTGGCATTAATGTTGACTTGGCAATTTCCACACGGCAAT	300
QY	301	ATCATCTACACAGCTTTTCTACGCTATGGAGAAATAATTTCTTGTGCTGAGGTG	360
DB	301	ATCATCTACACAGCTTTTCTACGCTATGGAGAAATAATTTCTTGTGCTGAGGTG	360
QY	361	TCAGAAAATTACTTCAAATAATAGGCATAGATAAGATTAAAGAAATTCATACACCATG	420
DB	361	TCAGAAAATTACTTCAAATAATAGGCATAGATAAGATTAAAGAAATTCATACACCATG	420
QY	421	ACCACGACCATCACTCAGACACAGCAGTCACTCAGACCATGAGCGTCACTCAGACCATG	480
DB	421	ACCACGACCATCACTCAGACACAGCAGTCACTCAGACCATGAGCGTCACTCAGACCATG	480
QY	481	AGCATCACTCAGACACAGCAGTCACTCAGACCATGATCATCTCACCATAATCATG	540
DB	481	AGCATCACTCAGACACAGCAGTCACTCAGACCATGATCATCTCACCATAATCATG	540
QY	541	CTCGTTCGTGTAATAATTAAGCCAAAGCTTTTGGCCAGACCATGACTCAGATGATTCAG	600
DB	523	CTCGTTCGTGTAATAATTAAGCCAAAGCTTTTGGCCAGACCATGACTCAGATGATTCAG	582
QY	601	CTAAGATCTCTAGAAACAGCAGCGGGAAGAGCTCACCGACCAAGACATGCCAGTGGTA	660
DB	583	GTAAGATCTCTAGAAACAGCAGCGGGAAGAGCTCACCGACCAAGACATGCCAGTGGTA	642
QY	661	GAAGGAATGTCAGGACAGTGTAGTCTGCTAGTGAAGTGAAGTCAACTGTGTACAACTG	720
DB	643	GAAGGAATGTCAGGACAGTGTAGTCTGCTAGTGAAGTGAAGTCAACTGTGTACAACTG	702
QY	721	TCCTCTGAAGGAACCTCACTTTCTAGAGACAATAGAGACTCCAAGACTCGAAAACCTTCC	780
DB	703	TCCTCTGAAGGAACCTCACTTTCTAGAGACAATAGAGACTCCAAGACTCGAAAACCTTCC	762
QY	781	CCAAAGATGTAGCAGCTCCACTCCACCAGTGTACATCAAAAGAGCGGGTACCGCGC	840
DB	763	CCAAAGATGTAGCAGCTCCACTCCACCAGTGTACATCAAAAGAGCGGGTACCGCGC	822
QY	841	TGCTCGTAGGAAAAACAAATGATCTGTGAGTGAGCCCGGAAAGGCTTTATGTAATCCA	900
DB	823	TGCTCGTAGGAAAAACAAATGATCTGTGAGTGAGCCCGGAAAGGCTTTATGTAATCCA	882
QY	901	GAACACAAATGAAATCTCTAGGAGTGTGTTCAATGCAATCAAGCTACTGACATCTCATG	960
DB	883	GAACACAAATGAAATCTCTAGGAGTGTGTTCAATGCAATCAAGCTACTGACATCTCATG	942
QY	961	GCATGGGCATCCAGGTTCCGCGTAATGCAACAGAGTTTCAACTATCTCTGTCAGGCCATCA	1020
DB	943	GCATGGGCATCCAGGTTCCGCGTAATGCAACAGAGTTTCAACTATCTCTGTCAGGCCATCA	1002
QY	1021	TCACCAAATGTGATGTAGATCTTGTCTGATTCATACAGTGAAAGAGGCTGAAATCC	1080
DB	1003	TCACCAAATGTGATGTAGATCTTGTCTGATTCATACAGTGAAAGAGGCTGAAATCC	1062

QY	1081	CTCAAAGACCTATTTCATTACAAATAGCTGGGTTGGTGGTCTTTATAGCCATTTCATCA	1140
DB	1063	CTCAAAGACCTATTTCATTACAAATAGCTGGGTTGGTGGTCTTTATAGCCATTTCATCA	1122
QY	1141	TCAGTTTCTGCTCTGCTGGGGGTATCTTAGTGCCCTCATGAATCGGGTGTTTTTCA	1200
DB	1123	TCAGTTTCTGCTCTGCTGGGGGTATCTTAGTGCCCTCATGAATCGGGTGTTTTTCA	1182
QY	1201	AAITTTCTGCTAGTTTCTGTTGGCACTGGCCGTTGGGACCTTTCAGTGGTGATGCTTTT	1260
DB	1183	AAITTTCTGCTAGTTTCTGTTGGCACTGGCCGTTGGGACCTTTCAGTGGTGATGCTTTT	1242
QY	1261	TACACCTTCTCCACATTCATGCAAGTCAACCACCATAGTCATAGCCATCAAGAACCAG	1320
DB	1243	TACACCTTCTCCACATTCATGCAAGTCAACCACCATAGTCATAGCCATCAAGAACCAG	1302
QY	1321	CAATGGAATGAAGAAGAGGACCACTTTTCAGTCATCTGCTTCTCAAAACATAGAAGAAA	1380
DB	1303	CAATGGAATGAAGAAGAGGACCACTTTTCAGTCATCTGCTTCTCAAAACATAGAAGAAA	1362
QY	1381	GTGCCATTTTTCATTCACAGTGGAAAGGCTCTAACAGCTCTAGGAGGCTGTATTTCAATG	1440
DB	1363	GTGCCATTTTTCATTCACAGTGGAAAGGCTCTAACAGCTCTAGGAGGCTGTATTTCAATG	1422
QY	1441	TTCTGTTGAACATGTCCTCACATGTGATCAACAAATTTAAAGATAAGAAAGAAAAGATC	1500
DB	1423	TTCTGTTGAACATGTCCTCACATGTGATCAACAAATTTAAAGATAAGAAAGAAAAGATC	1482
QY	1501	AGAGAAGCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1560
DB	1483	AGAGAAGCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1542
QY	1561	CTCAACTTTCAACAATGAGGAGAAAGTAGATACAGATGATCGAACTGAAGGCTATTTCAC	1620
DB	1543	CTCAACTTTCAACAATGAGGAGAAAGTAGATACAGATGATCGAACTGAAGGCTATTTCAC	1602
QY	1621	GAGCAGACTCAAGAGAGCCCTCCACTTTGATTTTCAGCAGCCCTGCACTCTCGGAAGAG	1680
DB	1603	GAGCAGACTCAAGAGAGCCCTCCACTTTGATTTTCAGCAGCCCTGCACTCTCGGAAGAG	1662
QY	1681	NAGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1740
DB	1663	NAGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1722
QY	1741	GGTGCAGGAATAAATGCCATTCACATTTTCACAGTACACTCGCCAGCTCAGACGATCTCA	1800
DB	1723	GGTGCAGGAATAAATGCCATTCACATTTTCACAGTACACTCGCCAGCTCAGACGATCTCA	1782
QY	1801	TTCAACACCATCATGACTACCATCATATTTCCCATCATCATCACACCAAAACACCAATC	1860
DB	1783	TTCAACACCATCATGACTACCATCATATTTCCCATCATCATCACACCAAAACACCAATC	1842
QY	1861	CTCAGACTCAGCCAGCCGCTACTCTCGGGAGGAGCTGAAAGATGCGCGGCTGCGCACTT	1920
DB	1843	CTCAGACTCAGCCAGCCGCTACTCTCGGGAGGAGCTGAAAGATGCGCGGCTGCGCACTT	1902
QY	1921	TGGCCTGGATGGTGATTAATGGGTGATGGCGTGCACATTTTCAGGGATGGCCTAGCAATTG	1980
DB	1903	TGGCCTGGATGGTGATTAATGGGTGATGGCGTGCACATTTTCAGGGATGGCCTAGCAATTG	1962
QY	1981	GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTAAAGTACTTCTGTTGCTGTGCTCTGTC	2040
DB	1963	GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTAAAGTACTTCTGTTGCTGTGCTCTGTC	2022
QY	2041	ATGAGTTGCCCTCATGAATTAGGTGACTTTTGGCTTCTACTAAAGGCTGGCATGACCGTTA	2100
DB	2023	ATGAGTTGCCCTCATGAATTAGGTGACTTTTGGCTTCTACTAAAGGCTGGCATGACCGTTA	2082
QY	2101	AGCAGGCTGCTCTTTATAAATGCATTTGTACGCCATGCTGGCGTATCTTGGAAATGCAACAG	2160
DB	2083	AGCAGGCTGCTCTTTATAAATGCATTTGTACGCCATGCTGGCGTATCTTGGAAATGCAACAG	2142
QY	2161	GAATTTTTCATTTGGTCAATTAACCTCAAAATCTTTCTATGTGGATATATTTGCATTACTGCTG	2220

[illegible]

QY	1628	CTCACAAGAGCCCTCCCACATTGATTTGATTTTCTCAGCAGCGCTGGAGTAAGAAGAGGT	1622
Db	1563	CTCNCAAGAGCCCTCCCACTTTTGATTTCTCAGCAGCGCTGGAGTAAGAAGAGGT	1747
QY	1688	CATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCAGAGGGTGCAA	1682
Db	1623	CATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCAGAGGGTGCAA	1807
QY	1748	GAATAAATGCCATTCACATTTCCACGATACACTCGGCAGGTCAGACGATCTCATCACCA	1742
Db	1683	GAATAAATGCCATTCACATTTCCACGATACACTCGGCAGGTCAGACGATCTCATCACCA	1867
QY	1808	CCATCATGACTACCATCATATCTCCATCATCATCACACACACAAACACCACTCTCACAG	1802
Db	1743	CCATCATGACTACCATCATATCTCCATCATCATCACACACAAACACCACTCTCACAG	1927
QY	1868	TCACAGCCAGCGTACTCTCGGGAGGAGCTGMAAGATGCGGGCGTGGCCACTTTGGCCGTG	1862
Db	1803	TCACAGCCAGCGTACTCTCGGGAGGAGCTGMAAGATGCGGGCGTGGCCACTTTGGCCGTG	1987
QY	1928	GATGCTGATATGGGTGATGGCTGCACAAATTCACGATGGCTAGCAATTTGGTGTCTGC	1922
Db	1863	GATGCTGATATGGGTGATGGCTGCACAAATTCACGATGGCTAGCAATTTGGTGTCTGC	2047
QY	1988	TTTTACTGAAGCTTATCAAGTGGTTTAAGTACTTCTGCTGTGTTCTGCTCATGAGTT	1982
Db	1923	TTTTACTGAAGCTTATCAAGTGGTTTAAGTACTTCTGCTGTGTTCTGCTCATGAGTT	2107
QY	2048	GCCTCATGAATTAGTGACTTTGCTTCTACTAAAGGCTGCATGACCGTTAAGCAGGC	2042
Db	1983	GCCTCATGAATTAGTGACTTTGCTTCTACTAAAGGCTGCATGACCGTTAAGCAGGC	2167
QY	2108	TGTCCTTTATAATGCATGTGCAGCCATCTGGCGTATCTTGGAAATGGCAACAGGAATTT	2102
Db	2043	TGTCCTTTATAATGCATGTGCAGCCATCTGGCGTATCTTGGAAATGGCAACAGGAATTT	2227
QY	2168	CATTGCTCATTTATGCTGAAAATGTTTCTATGTGGATATTTGCACTACTGCTGGCTTATT	2162
Db	2103	CATTGCTCATTTATGCTGAAAATGTTTCTATGTGGATATTTGCACTACTGCTGGCTTATT	2287
QY	2228	CATGTATGTGCTGTGGTTGATATGGTACCTGAAATGCTGCACAAATGATAGTAGACCA	2222
Db	2163	CATGTATGTGCTGTGGTTGATATGGTACCTGAAATGCTGCACAAATGATAGTAGACCA	2347
QY	2288	TGGATGTAGCCGCTGGGGTATTTCTTTTACAAATGCTGGGATGCTTTTCGGGTTTGG	2282
Db	2223	TGGATGTAGCCGCTGGGGTATTTCTTTTACAAATGCTGGGATGCTTTTCGGGTTTGG	2407
QY	2348	AATTATGTACTATATTTCCATATTTTCAACATAAAAACTGTTTCGTATAAATTTCTAGTT	2342
Db	2283	AATTATGTACTATATTTCCATATTTTCAACATAAAAACTGTTTCGTATAAATTTCTAGTT	2467
QY	2408	AAGGTTTAAATGCTAGATAGCTTAAAGAGTCTCATAGTTTCAGTAGGTCATAGGAGGA	2402
Db	2343	AAGGTTTAAATGCTAGATAGCTTAAAGAGTCTCATAGTTTCAGTAGGTCATAGGAGGA	2527
QY	2468	TGAGTTTGATGCTGTACTATGACGCGTTTAAAGTATAGTGGGTTTGTGATTTTGTATT	2462
Db	2403	TGAGTTTGATGCTGTACTATGACGCGTTTAAAGTATAGTGGGTTTGTGATTTTGTATT	2587
QY	2528	GAATATTGCTGTCTGCTACAAAGTCAGTTTAAAGTACGTTTAAATATTTAAGTATATCTA	2522
Db	2463	GAATATTGCTGTCTGCTACAAAGTCAGTTTAAAGTACGTTTAAATATTTAAGTATATCTA	2647
QY	2588	TCTTGGAGATAAATCTGTATGTGCAATTCACCGGTATACCGTTTATATATGTAACAA	2582
Db	2523	TCTTGGAGATAAATCTGTATGTGCAATTCACCGGTATACCGTTTATATATGTAACAA	2707
QY	2648	GAGATTGGCATGACATGTTCTGTATGTTTCAGGAAAAATGCTTTAATGCTTTTCAA	2642
Db	2583	GAGATTGGCATGACATGTTCTGTATGTTTCAGGAAAAATGCTTTAATGCTTTTCAA	2767
QY	2708	GAACATAACACAGTATTTCTCTATACATGGAATTTAGTCTCTGAAGAATCTCGGTGTTAG	

|||||
Db 2643 GAACTAACACACTTATCTCTACTGATTTTAGTCTCTGAAGAACTGCTGGTGTAG 2702
Qy 2768 GAATAGATGTCATGAAGCTTAAATACCAAGAAAGCTTATCTACTGAATTTAAGCAAG 2827
Db 2703 GAATAGATGTCATGAAGCTTAAATACCAAGAAAGCTTATCTACTGAATTTAAGCAAAA 2762
Qy 2828 AAATAAGCAGAAA 2841
Db 2763 AAAAAAAAAAAAAA 2776

RESULT 4
HSU41060
LOCUS
DEFINITION Homo sapiens estrogen regulated LIV-1 protein (LIV-1) mRNA, PRI 08-FEB-2001
complete cds.
ACCESSION U41060
VERSION U41060.2 GI:12711792
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2744)
Green, C. and Morgan, H.
Direct Submission
Submitted (08-FEB-2001) Biochemistry, University of Liverpool, P.O.
Box 147, Liverpool L69 3BX, UK
REFERENCE 2 (bases 1 to 2744)
Green, C., Gilhooly, E. M. and Walker, N. J.
Direct Submission
Submitted (21-NOV-1995) Biochemistry, University of Liverpool, P.O.
Box 147, Liverpool L69 3BX, UK
COMMENT On Feb 8, 2001 this sequence version replaced gi:1256000.
FEATURES
source
1..2744
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="MCF-7 human breast cancer cell line"
/note="estrogen induced mRNA"
1..2744
/gene="LIV-1"
138..2387
/gene="LIV-1"
/note="breast cancer"
/codon_start=1
/product="estrogen regulated LIV-1 protein"
/protein_id="AAA96258.2"
/db_xref="GI:12711793"
translation="MARLKVILITFALSVTNPLHKLAAAFPTOTKISPWESGI
NVDLAISTROYHLOOLFVYGENNSLVGEFKLLQNIQIDIKRIHIHDDHSDH
EHSOHERSHDHEHSDHSHDHAAGSKNKRKALCPDHDSDSGKDPNRSQCKGA
HRPEHASGRNRKVDSSASEVSTVNTVSEGHFLETETPRPKLPKDVSSSTPP
SVTSKSRVLAGRTNESVSPRGMYSRNTNENPQCFNASKLLTSHGMGIQVPL
NATEFNLYCPAIIQIDARSLIHTESEKAEIPPKTYSQIAWVGFTAISITISLS
LGVLYPLMNRVFEKLLSFLVALAVGTLSGDAFLHLLPHSHASHSHSHEEPAME
KCPLESHLSSONIESAYFSDTWKGLTALGLYFMFLVHVLTLIKQFKKKKNQK
KPNDDVEIKQLSKXESQLSSTNEEKVDTDRTTEGYLRADSOEPSHDSQOPAVLEE
EYVIAAHAPQYVNEYVPRGNKNCHSFHDTLGSDDLIHHHHDYHHLHHHHQON
HPPHSQSRYSRELDAGVATLWVIMGDLNPFSDGLAIGAAFTIGLSGLTSV
AVFCHPELHEDGLVLLKAGMTVKQAVLYNALSNLAFYLGMAFTIGLYAENYSMW
IFALTAGLEMYVALVDMVPEMLHNDASDHGCSRWGYFFLQNAQMLLGFIMLLISIE
HKIVFRNF"
BASE COUNT 786 a 585 c 601 g 772 t
ORIGIN
Query Match 78.5%; Score 2716; DB 9; Length 2744;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2744; Conservative 0; Mismatches 0; Indels 18; Gaps 1;
Qy 1 CTCGTGCCGAATTCGGCAGACAGCGGCTGTTCCGCCCTGGTAGAGATTCTCGAAGACA 60
|||||
Db 1 CTCGTGCCGAATTCGGCAGACAGCGGCTGTTCCGCCCTGGTAGAGATTCTCGAAGACA 60
Qy 61 CCAGTGGGCCCGGTGTGGAAACCAACCTGCGCGGTGGCGGGCGGTGGGACACAGAGGCC 120
Db 61 CCAGTGGGCCCGGTGTGGAAACCAACCTGCGCGGTGGCGGGCGGTGGGACACAGAGGCC 120
Qy 121 GCGGAGACGAGGCGCAATGCGAGGAAGTTATCTGTAAATCTTTGATCCTGACCTTTGCC 180
Db 121 GCGGAGACGAGGCGCAATGCGAGGAAGTTATCTGTAAATCTTTGATCCTGACCTTTGCC 180
Qy 181 TCTCTGTACAAAATCCCTTTCATGAACCTAAAGCAGCTGCTTTCCGCCAGACCATGAGA 240
Db 181 TCTCTGTACAAAATCCCTTTCATGAACCTAAAGCAGCTGCTTTCCGCCAGACCATGAGA 240
Qy 241 AATTAGTCGGAATGGGAATCTGGCAATTAATGTTGACTTGGCAATTTCCACAGGCAAT 300
Db 241 AATTAGTCGGAATGGGAATCTGGCAATTAATGTTGACTTGGCAATTTCCACAGGCAAT 300
Qy 301 ATCATCTACACAGCTTTTCTACCGCTATGGAGAAAATAATTTCTGTCAGTTGAAGGGT 360
Db 301 ATCATCTACACAGCTTTTCTACCGCTATGGAGAAAATAATTTCTGTCAGTTGAAGGGT 360
Qy 361 TCAGAAAATTACTTCAAAATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATG 420
Db 361 TCAGAAAATTACTTCAAAATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATG 420
Qy 421 ACCAGGACCATCTCAGACCACGAGCATCACTCAGACCATGAGCGCTCACTCAGACCATG 480
Db 421 ACCAGGACCATCTCAGACCACGAGCATCACTCAGACCATGAGCGCTCACTCAGACCATG 480
Qy 481 AGCATCACTCAGACCACGAGCATCACTCTGACCATGATCATCACTCTCACCATAATCATG 540
Db 481 AGCATCACTCAGACCACGAGCATCACTCTGACCATGATCATCACTCTCACCATAATCATG 540
Qy 541 CTGCTTCTGGTAAATTAAGGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAG 600
Db 541 CTGCTTCTGGTAAATTAAGGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAG 600
Qy 583 GTAAAGATCTCTAGAAACAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 642
Db 583 GTAAAGATCTCTAGAAACAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 642
Qy 661 GAAGGAATGTCAAGGACAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 720
Db 661 GAAGGAATGTCAAGGACAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 720
Qy 702 GAAGGAATGTCAAGGACAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 702
Db 702 GAAGGAATGTCAAGGACAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 702
Qy 721 TCTCTGAAGGAATCTCACTTTCTAGAGACAATAGAGACTCCAGAGCTGGAAGAACTCTTCC 780
Db 721 TCTCTGAAGGAATCTCACTTTCTAGAGACAATAGAGACTCCAGAGCTGGAAGAACTCTTCC 780
Qy 703 TCTCTGAAGGAATCTCACTTTCTAGAGACAATAGAGACTCCAGAGCTGGAAGAACTCTTCC 762
Db 703 TCTCTGAAGGAATCTCACTTTCTAGAGACAATAGAGACTCCAGAGCTGGAAGAACTCTTCC 762
Qy 781 CCAAGATGTAAAGCAGCTCCACTCCACCAGTGTCCACATCAAAGAGCGGGGTGAGCGCGC 840
Db 781 CCAAGATGTAAAGCAGCTCCACTCCACCAGTGTCCACATCAAAGAGCGGGGTGAGCGCGC 840
Qy 841 TGGCTGTAGGAAAACAAATGAATCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 900
Db 841 TGGCTGTAGGAAAACAAATGAATCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 900
Qy 882 TGGCTGTAGGAAAACAAATGAATCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 882
Db 882 TGGCTGTAGGAAAACAAATGAATCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 882
Qy 901 GAAACAAATGAAAATCTCAGGAGTGTTCATGTCATCAAAGCTACTGACATCTCATG 960
Db 901 GAAACAAATGAAAATCTCAGGAGTGTTCATGTCATCAAAGCTACTGACATCTCATG 960
Qy 942 GAAACAAATGAAAATCTCAGGAGTGTTCATGTCATCAAAGCTACTGACATCTCATG 942
Db 942 GAAACAAATGAAAATCTCAGGAGTGTTCATGTCATCAAAGCTACTGACATCTCATG 942
Qy 961 GCATGGGCATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTCTGCCAGGCATCA 1020
Db 961 GCATGGGCATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTCTGCCAGGCATCA 1020
Qy 1002 TCAACCAAAATTTGATGCTAGATCTTGTCTCATCAACAAGTGAAGAGAGGCTGAATCC 1080
Db 1002 TCAACCAAAATTTGATGCTAGATCTTGTCTCATCAACAAGTGAAGAGAGGCTGAATCC 1080
Qy 1081 CTCCTGAAGACCTATTCTTCAATACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGG 1140
Db 1081 CTCCTGAAGACCTATTCTTCAATACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGG 1140
Qy 1063 CTCCTGAAGACCTATTCTTCAATACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGG 1122
Db 1063 CTCCTGAAGACCTATTCTTCAATACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGG 1122

QY 1141 TCAGTTTCCCTGCTCTGCTGGGGTATCTTAGTGCCTCTCATGAATCGGGTGTCTTTTCA 1200
DB 1123 TCAGTTTCCCTGCTCTGCTGGGGTATCTTAGTGCCTCTCATGAATCGGGTGTCTTTTCA 1182
QY 1201 AATTTCCCTGAGTTTCTGCTGGCACTGGCCGTTGGGACTTTGAGTGGTATGCTTTT 1260
DB 1183 AATTTCCCTGAGTTTCTGCTGGCACTGGCCGTTGGGACTTTGAGTGGTATGCTTTT 1242
QY 1261 TACACCTTCTTCCACATTTCTATGCAAGTCACCACTATGCTATAGCCATGAAGAACACAG 1320
DB 1243 TACACCTTCTTCCACATTTCTATGCAAGTCACCACTATGCTATAGCCATGAAGAACACAG 1302
QY 1321 CAATGGAATGAAAGAGGACCACTTTTCACTCATCTGCTCTCAAAACATAGAAGAA 1380
DB 1303 CAATGGAATGAAAGAGGACCACTTTTCACTCATCTGCTCTCAAAACATAGAAGAA 1362
QY 1381 GTGCTATTTTGAATTCACACCTGGAAGGGTCTAACAGCTCTAGAGGCCCTGTATTTTCATGT 1440
DB 1363 GTGCTATTTTGAATTCACACCTGGAAGGGTCTAACAGCTCTAGAGGCCCTGTATTTTCATGT 1422
QY 1441 TTCTGTTGAACATGCTCTCACATTTGATCAACAAATTAAGATAGAAGAAAAGATC 1500
DB 1423 TTCTGTTGAACATGCTCTCACATTTGATCAACAAATTAAGATAGAAGAAAAGATC 1482
QY 1501 ACAAGAACCTGAAATCATGATCTGGAGATTAAGAAGCAGTTGTCCCAAGTATGAAT 1560
DB 1483 AGAAGAACCTGAAATGATGATGTGGAGATTAAGAGCAGTTGTCCCAAGTATGAAT 1542
QY 1561 CTCAACTTTCAACAAATGAGGAGAACTAGATACAGATGATGAACTGGAAGGCTATTTTAC 1620
DB 1543 CTCAACTTTCAACAAATGAGGAGAACTAGATACAGATGATGAACTGGAAGGCTATTTTAC 1602
QY 1621 GAGCAGCTCACAGAGCCCTCCACTTTGATTTCTCAGCAGCCTGCAAGTCTTGGAAGAAG 1680
DB 1603 GAGCAGCTCACAGAGCCCTCCACTTTGATTTCTCAGCAGCCTGCAAGTCTTGGAAGAAG 1662
QY 1681 AAGAGTCTATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAG 1740
DB 1663 AAGAGTCTATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAG 1722
QY 1741 GGTGCAAGAAATAAATGCCATTCATTTCCACGATACACTCGGCCGCTCAGACGATCTCA 1800
DB 1723 GGTGCAAGAAATAAATGCCATTCATTTCCACGATACACTCGGCCGCTCAGACGATCTCA 1782
QY 1801 TTCACACCATCATGACTACCATCATATTTCCCATCATCATCCACACCAACCAACCATC 1860
DB 1783 TTCACACCATCATGACTACCATCATATTTCCCATCATCATCCACACCAACCAACCATC 1842
QY 1861 CTCACAGTCACAGCAGCGCTACTCTCGGGAGAGCTGAAGATGCGCGGCTCGCCACTT 1920
DB 1843 CTCACAGTCACAGCAGCGCTACTCTCGGGAGAGCTGAAGATGCGCGGCTCGCCACTT 1902
QY 1921 TGGCTGATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980
DB 1903 TGGCTGATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1962
QY 1981 GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAGTACTTCTGTGCTGTGCTGTC 2040
DB 1963 GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAGTACTTCTGTGCTGTGCTGTC 2022
QY 2041 ATGAGTTGCCATCAATAGTCACTTTGCTGTTCTACTAAAGCTGGCATGACCGTTA 2100
DB 2023 ATGAGTTGCCATCAATAGTCACTTTGCTGTTCTACTAAAGCTGGCATGACCGTTA 2082
QY 2101 AGCAGCTGCTCTTTTATATGATGTCAGCAGCTGCGGCTATCTTGGAAATGCAACAG 2160
DB 2083 AGCAGCTGCTCTTTTATATGATGTCAGCAGCTGCGGCTATCTTGGAAATGCAACAG 2142
QY 2161 GAAATTTTCATTTGATGCTGAAATGTTTCTATGATGATATTTGCACTTTACTGCTG 2220
DB 2143 GAAATTTTCATTTGATGCTGAAATGTTTCTATGATGATATTTGCACTTTACTGCTG 2202

QY 2221 GCTTATTTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
DB 2203 GCTTATTTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2262
QY 2281 GTGACCATGGATGTAGCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG 2340
DB 2263 GTGACCATGGATGTAGCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG 2322
QY 2341 GTTTTGGAAATATGTTTACTTATTTCCATATTTGAACATATAAATCGTGTTCGTATAAAT 2400
DB 2323 GTTTTGGAAATATGTTTACTTATTTCCATATTTGAACATATAAATCGTGTTCGTATAAAT 2382
QY 2401 TCTAGTTAAGGTTTAAATGCTAGAGTAGCTTAAAGTGTGTCATAGTTTTCAGTAGGTCAT 2460
DB 2383 TCTAGTTAAGGTTTAAATGCTAGAGTAGCTTAAAGTGTGTCATAGTTTTCAGTAGGTCAT 2442
QY 2461 AGGAGATGAGTTTGTATGCTGCTACTATGTCAGCGTTTAAAGTGTAGTGGTGTGTTGATTT 2520
DB 2443 AGGAGATGAGTTTGTATGCTGCTACTATGTCAGCGTTTAAAGTGTAGTGGTGTGTTGATTT 2502
QY 2521 TTGTATTGAATATGCTGCTGCTTACAAAGTCAGTTAAAGGTACGTTTAAATATTTAAGT 2580
DB 2503 TTGTATTGAATATGCTGCTGCTTACAAAGTCAGTTAAAGGTACGTTTAAATATTTAAGT 2562
QY 2581 TATTCATCTTGGAGATAAAATCTGTATGCAATTCACCGCTATTACCAGTTTATTATG 2640
DB 2563 TATTCATCTTGGAGATAAAATCTGTATGCAATTCACCGCTATTACCAGTTTATTATG 2622
QY 2641 TAAACAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGGAAAAATGCTTTTAAATGCT 2700
DB 2623 TAAACAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGGAAAAATGCTTTTAAATGCT 2682
QY 2701 TTTTCAAGAACTAAACACAGTTTATCTATCTACTGGAATTTTAGCTCTCTGAAGAACTGCTGG 2760
DB 2683 TTTTCAAGAACTAAACACAGTTTATCTATCTACTGGAATTTTAGCTCTCTGAAGAACTGCTGG 2742
QY 2761 TG 2762
DB 2743 TG 2744

RESULT 5
176892 LOCUS
DEFINITION Sequence 3 from patent US 5693465.
ACCESSION I76892
VERSION I76892.1 GI:3013046
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2404)
AUTHORS Manning D.Lockwood., Nicholson R.Ian., Gee,J.Margaret.Wendy. and Green,C.Douglas.
TITLE Methods for predicting the behaviour of breast tumours
JOURNAL Patent: US 5693465-A 3 02-DEC-1997;
FEATURES
source Location/Qualifiers
i. .2404
/organism="unknown"
BASE COUNT 726 a 411 c 492 g 775 t
ORIGIN

Query Match 65.0%; Score 2249; DB 6; Length 2404;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 2341; Conservative 0; Mismatches 0; Indels 62; Gaps 3;

QY 1119 GGTTTTATAGCCATTTCCATCATCAGTTTCCCTGCTCTGCTGGGGTATCTTAGTGCCT 1178
DB 1 GGTTTTATAGCCATTTCCATCATCAGTTTCCCTGCTCTGCTGGGGTATCTTAGTGCCT 60
QY 1179 CTCATGAATCGGGTGTGTTTTTCAAAATTTCCCTGAGTTTCTGTCGCACTGGCCGTTGG 1238
DB 61 CTCATGAATCGGGTGTGTTTTTCAAAATTTCCCTGAGTTTCTGTCGCACTGGCCGTTGG 120

QY 1239 ACTTTGAGTGGTGATGCTTTTACACCTCTCCACATCTCTCATGCAAGTCACCCACAT 1298
DB 121 ACTTTGAGTGGTGATGCTTTTACACCTCTCTCCACATCTCTCATGCAAGTCACCCACAT 180
QY 1299 AGTCATAGCCATGAAGAACAGCAATGGAATGAAAGAGAGGAGCCACTTTTCAGTCATCTG 1358
DB 181 AGTCATAGCCATGAAGAACAGCAATGGAATGAAAGAGAGGAGCCACTTTTCAGTCATCTG 240
QY 1359 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTCATTTCCACGCTGGAGGGTCTTAACAGCT 1418
DB 241 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTCATTTCCACGCTGGAGGGTCTTAACAGCT 300
QY 1419 CTAGAGGCCCTGATTTTCATGTTTCTTTGTTGAACATGCTCCTCACATTTGATCAACAACATTT 1478
DB 301 CTAGAGGCCCTGATTTTCATGTTTCTTTGTTGAACATGCTCCTCACATTTGATCAACAACATTT 360
QY 1479 AAAGATGAAGAAAGAAAGATCAGAGAAACCTGAAATGATGATGATGTTGGAGATTAAG 1538
DB 361 AAAGATGAAGAAAGAAAGATCAGAGAAACCTGAAATGATGATGATGTTGGAGATTAAG 420
QY 1539 AAGCAGTTGTCCAACTATGAATCTCAACTTTCAACAAATGAGGAGAAAGTAGATACAGAT 1598
DB 421 AAGCAGTTGTCCAACTATGAATCTCAACTTTCAACAAATGAGGAGAAAGTAGATACAGAT 480
QY 1599 GATCGAACTGAAGCTATTTACGAGAGACTCACAAGAGCCCTCCCACTTTGATTTCTCAG 1658
DB 481 GATCGAACTGAAGCTATTTACGAGAGACTCACAAGAGCCCTCCCACTTTGATTTCTCAG 540
QY 1659 CAGCCTCAGCTCTGGAAGAGAGAGGTGATGATAGCTCATGCTCATCCACAGGAAAGTC 1718
DB 541 CAGCCTCAGCTCTGGAAGAGAGAGGTGATGATAGCTCATGCTCATCCACAGGAAAGTC 600
QY 1719 TACAATGAATATGTACCCAGAGGTCGAAGAATAAATGGCAATTCACATTTCCACGATACA 1778
DB 601 TACAATGAATATGTACCCAGAGGTCGAAGAATAAATGGCAATTCACATTTCCACGATACA 660
QY 1779 CTGGCCAGTCACAGCATCTCATCCACAGCATCATGATAGCTCATGCTCATCCACAGGAAAGTC 1838
DB 661 CTGGCCAGTCACAGCATCTCATCCACAGCATCATGATAGCTCATGCTCATCCACAGGAAAGTC 720
QY 1839 CACCACCAAAACCCACCATCTCACAGTCACAGCCAGCGTACTCTCGGAGGAGCTG 1898
DB 721 CACCACCAAAACCCACCATCTCACAGTCACAGCCAGCGTACTCTCGGAGGAGCTG 780
QY 1899 AAAGATGCGCGGCTGCCACTTTGGCCTGGATGGTGAATAATGGGTGATGGCCTGCACAAAT 1958
DB 781 AAAGATGCGCGGCTGCCACTTTGGCCTGGATGGTGAATAATGGGTGATGGCCTGCACAAAT 840
QY 1959 TTCAGCGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTGGTTAAAGT 2018
DB 841 TTCAGCGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTGGTTAAAGT 900
QY 2019 ACTTCGTGTGCTGTCTCATGATGGTGCCTCATGAATTAGTGACTTTGCTGTTCTTA 2078
DB 901 ACTTCGTGTGCTGTCTGTCTCATGATGGTGCCTCATGAATTAGTGACTTTGCTGTTCTTA 960
QY 2079 CTAAGGCTGGCATGACCGTTAAGCAGGCTGTCTTTTAAATGCAATGTCAGCCATGCTG 2138
DB 961 CTAAGGCTGGCATGACCGTTAAGCAGGCTGTCTTTTAAATGCAATGTCAGCCATGCTG 1020
QY 2139 GGGTATCTTGGAAATGGCAACAGGAATTTTCATTTGCTCATTTATGCTGAAATGTTTCTATG 2198
DB 1021 GGGTATCTTGGAAATGGCAACAGGAATTTTCATTTGCTCATTTATGCTGAAATGTTTCTATG 1080
QY 2199 TGGATATTGGCACTTACTGCTGGCTTATTCATGATGTTGCTGTTGATATGTAACCT 2258
DB 1081 TGGATATTGGCACTTACTGCTGGCTTATTCATGATGTTGCTGTTGATATGTAACCT 1140
QY 2259 GAAATGCTGCAAAATGATGCTAGTCACCATGATGATACCCCTGGGGGTATTTCTTTTA 2318
DB 1141 GAAATGCTGCAAAATGATGCTAGTCACCATGATGATGATACCCCTGGGGGTATTTCTTTTA 1200

QY 2319 CAGAACTGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTTGAACAT 2378
DB 1201 CAGAACTGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTTGAACAT 1259
QY 2379 AAAATCGTGTTCGTATAAATTTCTAGTTTAAAGTTTAAATGCTAGAGTAGCTTTAAAAAGT 2438
DB 1260 AAAATCGTGTTCGTATAAATTTCTAGTTTAAAGTTTAAATGCTAGAGTAGCTTTAAAAAGT 1318
QY 2439 TGTCTATGTTTTCAGTAGTTCATAGGAGATGAGTTTGTATGCTGTACTATGAGGCTTTTA 2498
DB 1319 TGTCTATGTTTTCAGTAGTTCATAGGAGATGAGTTTGTATGCTGTACTATGAGGCTTTTA 1378
QY 2499 AAGTTAGTGGGTTTTCGTGATTTTGTATGAAATATGCTGTCTGTACAAAGCTCAGTTAA 2558
DB 1379 AAGTTAGTGGGTTTTCGTGATTTTGTATGAAATATGCTGTCTGTACAAAGCTCAGTTAA 1438
QY 2559 AGCTAGCTTTTAAATATTTAAAGTTTATCTATCTTGGAGATAAAAATCTGTATGTCAATTTCA 2618
DB 1439 AGCTAGCTTTTAAATATTTAAAGTTTATCTATCTTGGAGATAAAAATCTGTATGTCAATTTCA 1498
QY 2619 CCGGTATTTACCAGTTTATTTATGCTAAACAGAGATTTGGCATGACATGTTCTGTATGTTTC 2678
DB 1499 CCGGTATTTACCAGTTTATTTATGCTAAACAGAGATTTGGCATGACATGTTCTGTATGTTTC 1558
QY 2679 AGGGAAAAATGCTTTTAAATGCTTTTCAAGAACTAACACAGTTTTCCTATCTACTGGATTT 2738
DB 1559 AGGGAAAAATGCTTTTAAATGCTTTTCAAGAACTAACACAGTTTTCCTATCTACTGGATTT 1618
QY 2739 TAGGTCCTGAAAGAACTGCTGCTGTTTAAAGAAATAAGAAATGTCATGAAGCCTTAAATACC 2798
DB 1619 TAGGTCCTGAAAGAACTGCTGCTGTTTAAAGAAATAAGAAATGTCATGAAGCCTTAAATACC 1678
QY 2799 AAGAAAGCTTATACGAAATTTAAGCAAGAAATTAAGGAGAAAGAGAGAAATCTGAGAA 2858
DB 1679 AAGAAAGCTTATACGAAATTTAAGCAAGAAATTAAGGAGAAAGAGAGAAATCTGAGAA 1738
QY 2859 TTGGGAGGCACTAGATCTTATAAAATCACAATAATTTGTTGTAATTTAGAGGGGAGAA 2918
DB 1739 TTGGGAGGCACTAGATCTTATAAAATCACAATAATTTGTTGTAATTTAGAGGGGAGAA 1798
QY 2919 TTTAGAAATTAAGTATAAAAGGCAAGATTTAGTATAGCTACATTTCAATTTAAACATTTTGT 2978
DB 1799 TTTAGAAATTAAGTATAAAAGGCAAGATTTAGTATAGCTACATTTCAATTTAAACATTTTGT 1858
QY 2979 CAGATTTATTTCCCGTAAAAAGCTAGTGAGCACTCTCATATCTAAATTTAGTGTACATTTA 3038
DB 1859 CAGATTTATTTCCCGTAAAAAGCTAGTGAGCACTCTCATATCTAAATTTAGTGTACATTTA 1918
QY 3039 ACTTTGTATAATACAGAAATCTAAATATTTAATGAATTCAGCAATATACACTTGCAC 3098
DB 1919 ACTTTGTATAATACAGAAATCTAAATATTTAATGAATTCAGCAATATACACTTGCAC 1978
QY 3099 AAGAAATTTGAAATTTCAAAATGTTGCGGGTTATATACAGATGAGTACAGTGTAG 3158
DB 1979 AAGAAATTTGAAATTTCAAAATGTTGCGGGTTATATACAGATGAGTACAGTGTAG 2038
QY 3159 TTTATGATACACAGCTGGGTTATTCGCAAGTTATATATACCCAAAGCTGTATGACTG 3218
DB 2039 TTTATGATACACAGCTGGGTTATTCGCAAGTTATATATACCCAAAGCTGTATGACTG 2098
QY 3219 GATGTTCTGGTTACCTGGTTTACAAAATTTACAGAGTAGT----- 3258
DB 2099 GATGTTCTGGTTACCTGGTTTACAAAATTTACAGAGTAGTAAAACTTTGATATATATGAG 2158
QY 3259 ----- 3258
DB 2159 GATATTTAAACTACACTAAGTATATCTGATTTGATTCGATTTAGAAAATTTGATATATGAG 2218
QY 3279 GATATTTAAACTACACTAAGTATATCTGATTTGATTTGATTTGATATCTCTC 3338
DB 2219 GATATTTAAACTACACTAAGTATATCTGATTTGATTTGATTTGATATCTCTC 2278
QY 3339 AGTGCTTCAGTGCTATCATTTGTGAGCAATTTGCTTTTATATACGGTACTGTAGCCATCTA 3398

Db	2279	CTCAGTGCCTATCATCTTGACCAATTGCTCTTTATATACGGTACTGTAGCCATACTA	2338
Qy	3399	GGCCTGTCTGTGGCATTCTCTAGATGTTCTCTTTTTACACAATAAAATTCCTTATATACAGC	3458
Db	2339	GGCCNGTCTGTGGCATTCTCTAGATGTTCTCTTTTTACACAATAAAATTCCTTATATACAGC	2398
Qy	3459	TTG 3461	
Db	2399	TTG 2401	
LOCUS	BC008317		
DEFINITION	Homo sapiens, clone IMAGE:3343159, mRNA, partial cds.		
VERSION	BC008317		
KEYWORDS	BC008317.1 GI:14249878		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2236)		
JOURNAL	Strausberg, R.		
	Direct Submission		
	Submitted (25-MAY-2001) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk		
	Email: gcapbs@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Rubin Laboratory		
	DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Genome Sequence Centre,		
	BC Cancer Agency, Vancouver, BC, Canada		
	info@bcsc.bc.ca		
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,		
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,		
	Leticia Hsiao, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo		
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven		
	Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saedi, Jacqueline		
	Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,		
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,		
	George Yang, Scott Zuyderduyn, Marco Marra.		
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Series: IRAL Plate: 22 Row: k Column: 4.		
FEATURES	Location/Qualifiers		
Source	1..2236		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3343159"		
	/tissue_type="Eye, retinoblastoma"		
	/clone_lib="NIH_MGC_16"		
	/lab_host="DH10B-R"		
	/note=Vector: pOTB7"		
CDS	<1..1151		
	/codon_start=3		
	/product="Unknown (protein for IMAGE:3343159)"		
	/protein_id="AAH08317.1"		
	/db_xref="GI:14249879"		
	/translation="FLALLPHSHASHSHSHEEPAMEMKRGPLFSLSSQNIESAY		
	FDSWKGUAGLGLYFMFLVHLVTLIKFQKKKKKKPPENDDDVEIKKOLSKYES		
	QLSTNEBKVDTDTEGYLRADSGSFSDPVLSEEVNIAHAHPQEVNEYVP		
	RCKNKCHSHFDHTLGOSDLDIHHHDYHIILHHHHQHHPHSORYSRLELKDV		
	VATLAWVIINGDLHNFSGLGIGAATFEGLSGLSTSVAVFCELPHELGDFAVLK		
	AGNTVKAVLYNALSNLAYLGNATGIFIGHTAEVNSWMIFALTAGLMVALVDMP		
	EMLNDRASDHGCRWGTFFFLHGFGIMLLISIFEHKIVFRNF"		
BASE.COUNT	704 a 375 c 449 g 708 t		


```
QY 2275 ATGCTAGTACCATGGATGTAGCCGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGC 2334
DB 1021 ATGCTAGTACCATGGATGTAGCCGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGC 1080
QY 2335 TTTTGGGTTTGGAAATATGTTACTATTTCCATATTTGAACATAAAATCGTGTTCGTA 2394
DB 1081 TTTTGGGTTTGGAAATATGTTACTATTTCCATATTTGAACATAAAATCGTGTTCGTA 1140
QY 2395 TAAATTTCTAGTTAAGGTTTAAATGCTAGATAGCTTAAAAAGTTGTCTATAGTTTCAGTA 2454
DB 1141 TAAATTTCTAGTTAAGGTTTAAATGCTAGATAGCTTAAAAAGTTGTCTATAGTTTCAGTA 1200
QY 2455 GGTCTAGGGAGATGAGTTTGTCTGCTACTATGTCACGGTTTAAAGTTAGTGGGTTTG 2514
DB 1201 GGTCTAGGGAGATGAGTTTGTCTGCTACTATGTCACGGTTTAAAGTTAGTGGGTTTG 1260
QY 2515 TGAATTTTGTATGTAATTTGCTGCTGTCTTACAAAGTCAGTTAAAGGTACGTTTAAATAT 2574
DB 1261 TGAATTTTGTATGTAATTTGCTGCTGTCTTACAAAGTCAGTTAAAGGTACGTTTAAATAT 1320
QY 2575 TTAAGTTATTTCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACAGTTT 2634
DB 1321 TTAAGTTATTTCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACAGTTT 1380
QY 2635 ATTATGTAACAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGGAATAATGCTTTT 2694
DB 1381 ATTATGTAACAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGGAATAATGCTTTT 1440
QY 2695 AATGCTTTTTCAGAACTAACACAGTTATTCCTATCTACTGGATTTTAGTGTCTCTGAAGAC 2754
DB 1441 AATGCTTTTTCAGAACTAACACAGTTATTCCTATCTACTGGATTTTAGTGTCTCTGAAGAC 1500
QY 2755 TGTGTGTTTGTAGGAATAGATGTCGATGCAAGCTTAAATACCAAGAACTTATACGTG 2814
DB 1501 TGTGTGTTTGTAGGAATAGATGTCGATGCAAGCTTAAATACCAAGAACTTATACGTG 1560
QY 2815 AATTTAGCAAGAAATAAGGAGAGAAAGAGAAATCTGAGAAATTTGGGAGGCATAGAT 2874
DB 1561 AATTTAGCAAGAAATAAGGAGAGAAAGAGAAATCTGAGAAATTTGGGAGGCATAGAT 1620
QY 2875 TCTTATAAATCACAAAATTTGTTGTAATTTAGGAGGAGAAATTTAGNATTAAGTATA 2934
DB 1621 TCTTATAAATCACAAAATTTGTTGTAATTTAGGAGGAGAAATTTAGNATTAAGTATA 1680
QY 2935 AAAAGGCAGAAATAGTATAGATACATTTCAATTAACAAATTTTCTCAGGATTTATTCGGT 2994
DB 1681 AAAAGGCAGAAATAGTATAGATACATTTCAATTAACAAATTTTCTCAGGATTTATTCGGT 1740
QY 2995 AAAAACGTAGTGGACAC-TCTCATATATAATT--AGTGACATTTAACTTTGTATATA 3051
DB 1741 AAAAACGTAGTGGACAC-TCTCATATATAATT--AGTGACATTTAACTTTGTATATA 1800
QY 3052 CAGAAATCTAATATATTTATGATTAAGTCAAGCAATATA-CACTTGACCAAGAAATTCGAA 3110
DB 1801 CAGAAATCTAATATATTTATGATTAAGTCAAGCAATATACTCTGACCAAGAAATTCGAA 1860
QY 3111 TTTCAAAATGTTGCTGCGGGTTATATACCAGATGAGTACAGTGAAGTAG-TTTATGATCA 3169
DB 1861 TTTCAAAATGTTGCTGCGGG-TATATACCAGATGAGTACAGTGAAGTAGTTTATGATCA 1919
QY 3170 CCAGACTGGGTTATGCCAAGTTATATATACCAAAAGCTGTATGACTGGATGTTCTGGT 3229
DB 1920 CCAGACTGGGTTATGCCAAGTTATATATACCAAAAGCTGTATGACTGGATGTTCTGGT 1979
QY 3230 TACCTGGTTTACAAAATATCAGACTAGTAAACCTTTGATATATATGAGGATATTAAC 3289
DB 1980 TACCTGGTTTACAAAATATCAGACTAGTAAACCTTTGATATATATGAGGATATTAAC 2039
QY 3290 TACACTAAGTATCATTTGATTCGATTACAGAAAGTACTTTGATATCTCTCAGTGTTCAGT 3349
DB 2040 TACACTAAGTATCATTTGATTCGATTACAGAAAGTACTTTGATATCTCTCAGTGTTCAGT 2099
```

```
QY 3350 GCTATCATTTGTAGCAATTTGTC-TTTATATACGGTACTGTAGCCATACTAGGCTGTCTG 3408
DB 2100 GCTATCATTTGTAGCAATTTGTC-TTTATATACGGTACTGTAGCCATACTAGGCTGTCTG 2159
QY 3409 TGGCATTTCTAGATGTTTCTTTTACACAATAAATTCCTTATATACAGCTTG 3461
DB 2160 TGGCATTTCTAGATGTTTCTTTTACACAATAAATTCCTTATATACAGCTTG 2212

RESULT 7
AB071697 3287 bp mRNA linear ROD 15-MAR-2002
LOCUS Mus musculus ermelin mRNA for endoplasmic reticulum membrane
DEFINITION protein, complete cds.
ACCESSION AB071697
VERSION AB071697.1 GI:19570345
KEYWORDS Mus musculus
SOURCE Mus musculus skeletal muscle myoblasts cell_line:C2 cDNA to mRNA,
clone_lib:lambda ZAPII C2 myoblast cDNA.
ORGANISM Mus musculus
REFERENCE 1 Suzuki,A. and Endo,T.
AUTHORS Ermelin, an endoplasmic reticulum transmembrane protein, contains
TITLE the novel HELP domain conserved in eukaryotes
JOURNAL Gene 284 (1-2), 31-40 (2002)
MEDLINE 21888618
REFERENCE 2 (bases 1 to 3287)
AUTHORS Endo,T.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) Takeshi Endo, Chiba University, Department
of Biology, Faculty of Science; 1-33 Yayoi-cho, Inage-ku, Chiba,
Chiba 263-8522, Japan (E-mail: tendo@cuphd.nd.chiba-u.ac.jp,
Tel:81-43-290-3911, Fax:81-43-290-3911)
FEATURES
source
1..3287
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="C2"
/cell_type="skeletal muscle myoblasts"
/clone_lib="lambda ZAPII C2 myoblast cDNA"
1..3287
/gene="ermelin"
704..2221
/gene="ermelin"
/codon_start=1
/product="endoplasmic reticulum membrane protein"
/protein_id="BAB86300.1"
/db_xref="GI:19570346"
/translation="MYSRNTNDNIQECFNTTKLLTSHGMSIQALLNATPEFNYLCPAII
NOIDARACLIHTASEKKAIEIPKYSIQIANLGGFIAISISFLSLQVILVPLMNRV
FFKFLSLFVLAVGTLGSDALLHLPHSHASHSHSHSHEEPAMEMKRGPLFSLSAO
NTEESYFDSTWTKGLTGLGGLFMEFLVHVJTLKQFKDKRKKKKKKKEDVEDVESKK
QLSKYDSQLSNEEKVDPGERPSYLRADSPQSPFDDQPTMLEEEVEMIAHAHQPE
VYNEVPRGCKNKSHFDHTLGQSDDLIHHDHVDHHTLHHHHHONHHPHSHQRYSR
BELKDAGTATLAWVIMGDLGNFSDGLAIGAAFTGSLGSLSTSVAFHLPHELG
DFAVLLKAGMTYKQAVLYNALSAMLYLGMATGIFIGHYAEVNSHWIATLGLPFMYV
ALVDVWPEMLHNDASDHGCSRWCYFFLQNGILLGFGILLISIFEKIVFRINP"
BASE COUNT 931 a 731 c 751 g 874 t
ORIGIN
Query Match 50.7%; Score 1753.2; DB 10; Length 3287;
Best Local Similarity 75.5%; Pred. No. 0;
Matches 2506; Conservative 0; Mismatches 698; Indels 114; Gaps 22;
QY 196 CCCTTCATGAACATAA---AGCAGCTGCTTTCCCCAGACCACCTGAGAAAATTAGTCCGA 252
DB 1 CCCTTCATGAACATAAATCAACAGCTGCTTTCTCTCAGACTACTGAGAAAATAAATTCAA 60
QY 253 ATTGGGAATCTGGCAATTAATGTTCACTTGGCAATTTCCACACGCGCATATATCTCAAC 312
DB 61 ATTGGGAACCTGGTGTAAATGTTGACTTGGCAGTTACCATGCGACACACCATCTGCAGC 120
```

[illegible]

QY	1381	GTGCGCTATT	TGATTTC	CACGTTGG	AAGGCT	TAACAG	CTTAGAG	GCCTGT	ATTTCAT	GTG	1440
Db	1197	GCTCCTATT	TTTTGATT	TCCACG	TGGAAAG	CTGTG	ACGCTCT	TAGGGG	CTTAT	TTTTTCAT	GTG
QY	1441	TTCTTTGTT	GAACATG	CTCCT	CACATTG	ATCAAC	CAATTTAA	AGAT	TAGA	AGAAAAA	GAATC
Db	1257	TTCTTTG	TGGACACG	TACT	CACACT	GATC	CAAGCA	ATTTAA	AGAT	TAGAAAAA	GAAGAATC
QY	1501	AGAAGAA	ACCTTGA	AAATG	ATGATG	TATG	GGAGATT	AAG	AAGCAG	TGTTGCC	AAATGATGAAT
Db	1317	AAAGAAA	ACCTTGA	AAATG	ATGATG	TATG	GGAGATT	AAG	AAGCAG	TGTTGCC	AAATGATGAAT
QY	1561	CTCAACT	TTTCAACA	ATGAG	GAGAA	AGATAG	ATAC	AGATG	ATG	CAACTG	AGAGCTATTTAC
Db	1377	CTCAGCT	TTTCTCAAT	TGA	AGAAAG	TGAC	CGCCGGA	AGCAGCT	G	AAAGCT	ATCTGCG
QY	1621	GACGAG	ACTCA	AAAG	CCCTCC	CCACCTT	GATTTCT	CAG	ACGCTG	CGAGTCT	TGGAAAGAG
Db	1437	GAGCGAG	CTCCA	AGAG	CCCTCC	CCCTT	GATTTCC	CAG	ACGCGAG	GATTTGG	NAGAGG
QY	1681	AAGAGG	TCATG	ATAG	CTCATG	CTCCAC	AGGAAG	CTT	PACA	ATGAATG	TATGACCAGAG
Db	1497	AAGAGG	TCATG	ATAG	CCCATG	CCACAC	CCCAAG	AGTCT	PACA	ATGAATG	TATGACCAGAG
QY	1741	GGTCAGA	GAATP	AAATG	CGCATTC	CACATTTCC	ACAG	ATAC	CTCGG	CAGTAC	AGACAGATCTCA
Db	1557	GCTGCA	AGAACA	AGTGC	CACTTCA	CACTTCC	ACAG	TACG	TGG	CCACG	TCCGACGACCTCA
QY	1801	TTCA	CCACC	ATCATG	ACTAC	CTCATAT	TTCTCC	ATCATG	CA	CCACCA	CAAAACCA
Db	1617	TCCAC	ACCATCAT	AGAG	CTACCAT	CCATCACA	TTCTG	CA	CCAC	CA	CCACCA
QY	1861	CTCAG	GTCA	CAGC	CGCGT	ACTCTC	GGAGG	AGCTG	AAAG	ATCG	CGCGCTCGGCACTT
Db	1677	CTCAG	CCAC	AGCC	ACGGCT	ACTCTC	GGAGG	AGCTG	AAAG	ATCG	CGCGCTCGGCACTT
QY	1921	TGGC	CTGG	ATG	GTGAT	AAATGG	GTG	GGCTGC	CA	CAATTT	TTCAGCGATGGCCTAGCAATTG
Db	1737	TGGC	CTGG	ATG	GTGAT	AAATGG	GTG	GGCTGC	CA	CAATTT	TTCAGCGATGGCCTAGCAATTG
QY	1981	GTGCTG	CTTTTACT	G	AAGGCTT	ATCAAG	TGGTTTAA	GTACT	CTCTG	CTTCTG	CTGTTCTGTC
Db	1797	GTGCTG	CTTTTACT	G	AAGGCTT	ATCAAG	TGGTTTAA	GTACT	CTCTG	CTTCTG	CTGTTCTGTC
QY	2041	ATGAG	TTGC	CTCATG	ATAG	TGACT	TTGCTT	CTTACT	AAAG	GC	TGGCATGACCGTTA
Db	1857	ATGAG	TTGC	CTCATG	ATAG	TGACT	TTGCTT	CTTACT	AAAG	GC	TGGCATGACCGTTA
QY	2101	AGCAG	CGTG	CTCTT	ATATG	CTCTG	CA	CCATGTTGG	CCCTAC	TG	GGAAATGGCAACAG
Db	1917	AGCAG	CGTG	CTCTT	ATATG	CTCTG	CA	CCATGTTGG	CCCTAC	TG	GGAAATGGCAACAG
QY	2161	GAATTT	TCAT	TGTC	ATP	TGCTG	AAAAATG	TTTCTAT	TG	GTGAT	TATTTGCACTTCTG
Db	1977	GGAT	TATTC	ATCGGG	CACTT	ATG	CAGAAAATG	TTTCTAT	TG	GTGAT	TATTTGCACTTCTG
QY	2221	GCTT	ATTTCA	TGATG	TGCTT	CGTTG	TATG	TG	TATG	TG	TATG
Db	2037	GCTT	ATTTCA	TGATG	TGCTT	CGTTG	TATG	TG	TATG	TG	TATG
QY	2281	GTG	ACCAT	TGATG	TAG	CCGCTGG	GGGTATTTCT	TTTTAC	AG	AAATGCT	TGGGATGCTTTGG
Db	2097	GTGAT	CA	CGGAT	CTG	CCGCTGG	GGGTATTTCT	TTTTAC	AG	AAATGCT	TGGGATGCTTTGG
QY	2341	GT	TTTTCGA	ATTTATG	T	A	TTTATTT	TCATATTT	TGA	ACATA	TAATAATCGTGTTCGAT
Db	2157	GT	TTTTCGA	ATTTATG	T	A	TTTATTT	TCATATTT	TGA	ACATA	TAATAATCGTGTTCGAT
QY	2401	TCT	AGTT	T	AAAGG	T	CTAG	T	CTAG	T	CTAG
Db	2217	TCT	AGTT	T	AAAGG	T	CTAG	T	CTAG	T	CTAG
QY	2461	AG	-----	GG	AG	ATG	AGT	TTG	TATG	CTG	TACTATG
Db	2276	AG	-----	GG	AG	ATG	AGT	TTG	TATG	CTG	TACTATG

Db 2277 AGATGGTAGAAATACGTTTCATATGCTGTGATACACAGCATTTAAACTAGTGGTTTGC 2336
QY 2516 GATTTTGTATCAATATGCTCTCTGTT-----ACAAGTCAGT 2555
Db 2337 AATTTTCTCATCAATTTTGGCCATTTGTTATGCTTATAAGTCAGTTCATTCGGGTAAACA 2396
QY 2556 TAAAGGTACGTTTAAATATTAAGTTATCTCTCTGGAGATAAAATCTGTATGCAAT 2615
Db 2397 TAAAGGTACGTTTAAATATTAAGTTATCTCTCTGGAGATAAAATCTGTATGCAAT 2456
QY 2616 TCACCCGGTATTACCACTTTATTTATGTAACAAGAGATTGGCATGACATGTTCTGTATGT 2675
Db 2457 CCACCACTATTACCCAGCTGT-----ATGTTTCGGCATGCCGTGCTCTCTAT 2505
QY 2676 TTCAGGGAATAATGCTTTAAATGCTTTTCAAGAACTACACAGATTATTCCTATACGGA 2735
Db 2506 TCTAGAG- AATATGCTTTAGTCTCTCTCTGTAGCTTTAGCATTTACTTCCCTGCTGGA 2564
QY 2736 TTTTAGTCTCTGAAGAACTGCTGTGTTTAGGAATAAGATGTGCATGAAGCCCTAAAT 2795
Db 2565 GTTGGGCCCT-----TGCTCGTGTGAGAAAGATGGSCACAAACCCAGAC 2612
QY 2796 ACCAAGAAAGCTTATACGAAATTAAGCAAAAGAAATAAGGAGAAAGAGAAATCTGA 2855
Db 2613 ACCAAGCTTACACTGAGCAT-----AGTCAGGAAATAGGAGAAAGAGAACTCGGAG 2668
QY 2856 GAATTTGGGAGGATPAGATCTTTATAAAATCACAATAATTTGTTAAATAGAG-GGGA 2914
Db 2669 CATTTGAGGAGGTATATATGCTTTTGAATAATCACAAGGTGCTGTTGAACGTTGAGAGGG 2728
QY 2915 GAAATTTAGAAATTAAGTATAAAGGCGAATTTAGTATAGTACATTCATTAACATTT 2974
Db 2729 GAAATTTAGATGTTTAAAGGGGCAATGTTTATGTCAGAGTACGTCGCTGGACATCTG 2788
QY 2975 TTGTGAGGATTTATTCGGTAAAAACGTTAGTGACACTCTCATATACTAATAGTGACA 3034
Db 2789 TCAAAATCTTATGTTCCCTGTGAACA--AGTGAACACTCTGTATACAC---TTCATCTGCT 2843
QY 3035 TTTAACTTTGATATAACAGAAATCTAAATATATTTAAATGAATTCAGCAATATACACTT 3094
Db 2844 TGTACATCTCTGATGATAGAAATGTAATATGTT-----AATCCAAGCAATGTCACCT 2898
QY 3095 GACCAAGAAATTTGAAATTTCAAAATGTTCTGGGGGTTATATACCAGATCAGTACAGTA 3154
Db 2899 TGGCTGAGAAATGAGGCATTTAAATCTGTGTGG-----AGATGGGCCAAATGA 2948
QY 3155 GTAGTTTATGATACAC---AGACTGGGTTATTGCCAAGTTATATATACCAAAAGCTGT 3211
Db 2949 GTAACTAGTGTGTACACAGAGACTTGGTTTTCACCATGCTCTCTATGCCGCAAGTTAT 3008
QY 3212 ATGACTGGATGTTCTGTTACCTGGTTTACAAAATTTATCAGAGTAGTAAACTTTGATAT 3271
Db 3009 GTGA-----ACATTTCTAGTGTCTAGGCTTCCAAAGT--GGCAAGTATTAAACTTTAGCCC 3062
QY 3272 ATATGAGGATPATAAAC-----TACACTAAGTATCATTTGATTCGATTTACAGAACTACT 3326
Db 3063 ATGAGAGGATGTTAAGACCTGTTCTCAAAATATTTTGACTTCTTGGATTTAAAAAGCACT 3122
QY 3327 TGCATATCTCTAGTCTCAGTGCTA-----TCATTTGAGCAAAATGTC--TTTATATA 3379
Db 3123 TAGCTCTCTTCTCTCTCAGTCTGTTTAAAAATGTCATTTGTAACAACACTGCTCTTGCATGA 3182
QY 3380 CGGTACTGTAGGCATACTAGGCTGTCTGTGGCATCTCTCTAGATCTTCTTTTACACA 3439
Db 3183 TGACACTGTAGATACATTTGGGCTGTCTGACACTGTCTGATGTTTCTTTTAAAAA 3242
QY 3440 ATAAATTCCTTATATCAG 3457
Db 3243 ATTAAGTGTCAACCCAG 3260

RESULT 8

AX017261
LOCUS AX017261
DEFINITION Sequence 12 from Patent W09947669.
ACCESSION AX017261
VERSION AX017261.1 GI:10042179
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1597)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from tissue of breast tumors
JOURNAL Patent: WO 9947669-A 12 23-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source Location/Qualifiers
1..1597
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 476 a 229 c 337 g 555 t
ORIGIN
Query Match 43.0%; Score 1486.8; DB 6; Length 1597;
Best local similarity 99.0%; Pred. No. 0;
Matches 1559; Conservative 0; Mismatches 7; Indels 8; Gaps 6;
QY 1895 GCTGAAGATGCCGGCTGCCACTTTGGCTGGATGGTGATAAATGGCTGATGGCCTGCA 1954
Db 19 GCTGAAGATGCCGGCTGCCACTCTGGCTGGATGGTGATAAATGGCTGCA 78
QY 1955 CAATTTACAGGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGCTTATCAAGTGGTTT 2014
Db 79 CAATTTACAGGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGCTTATCAAGTGGTTT 138
QY 2015 AAGTACTTCTGTGCTGTCTGTCATCAGTTGGCTCATGAATTAGTGACTTTGCTGT 2074
Db 139 AAGTACTTCTGTGCTGTCTGTCATCAGTTGGCTCATGAATTAGTGACTTTGCTGT 198
QY 2075 TCTACTAAAGGCTGGCATGACCGTTAAGCAGCTGTCTTTTATATGCAATTTCTACAGCAT 2134
Db 199 TCTACTAAAGGCTGGCATGACCGTTAAGCAGCTGTCTTTTATATGCAATTTCTACAGCAT 258
QY 2135 GCTGGGATCTTGAATGGCAACAGGAATTTTCATTTGCTCATTTATGCTGAAAATGTTTC 2194
Db 259 GCTGGGATCTTGAATGGCAACAGGAATTTTCATTTGCTCATTTATGCTGAAAATGTTTC 318
QY 2195 TATGTGGATATTGCACTTACTGCTGGCTTATTCATGATGTTGCTCTGGTGTATGCT 2254
Db 319 TATGTGGATATTGCACTTACTGCTGGCTTATTCATGATGTTGCTCTGGTGTATGCT 378
QY 2255 ACCTGAAATGCTGCAATGATGCTAGTACCATGGATGGATGCTAGCCGCTGGGGTATTTCTT 2314
Db 379 ACCTGAAATGCTGCAATGATGCTAGTACCATGGATGGATGCTAGCCGCTGGGGTATTTCTT 438
QY 2315 TTTACAGAACTGCTGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTTGA 2374
Db 439 TTTACAGAACTGCTGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTTGA 498
QY 2375 ACATAAAATCGTGTTCCTGATAAAATTTCTAGTAAAGCTTTAAATGCTAGAGTACCTAAA 2434
Db 499 ACATAAAATCGTGTTCCTGATAAAATTTCTAGTAAAGCTTTAAATGCTAGAGTACCTAAA 558
QY 2435 AAGTTGTCATAGTTTTCAGTAGGTCATAGGGAGATGAGTTTGTATGCTGTACTATGACAGC 2494
Db 559 AAGTTGTCATAGTTTTCAGTAGGTCATAGGGAGATGAGTTTGTATGCTGTACTATGACAGC 618
QY 2495 TTTAAAGTTAGTGGGTTTCTGATTTTGTATTTGAATATTTGCTGTCTGTGTACAAAGTCAG 2554
Db 619 TTTAAAGTTAGTGGGTTTCTGATTTTGTATTTGAATATTTGCTGTCTGTGTACAAAGTCAG 678

Query Match		37.2%	Score 1288;	DB 6;	Length 1310;
Best Local Similarity		99.8%	Pred. No. 3.5e-279;		
Matches 1310;	Conservative	0;	Mismatches	0;	Indels
2;	Gaps				
QY	1119	GGTTTTATAGCCATTTCCCATCATCAGTTTCCTGTCCTGCTCTGCTGGGGTATCTTACGTGCCT	1178		
Db	1	GGTTTTATAGCCATTTCCCATCATCAGTTTCCTGTCCTGCTGGGGTATCTTACGTGCCT	60		
QY	1179	CTCATGAATCGGGTGTTTTTCAAATTTCTCCTGAGTTTCTTGTGTGCACTGGCGGTGGG	1238		
Db	61	CTCATGAATCGGGTGTTTTTCAAATTTCTCCTGAGTTTCTTGTGTGCACTGGCGGTGGG	120		
QY	1239	ACTTTGAGTGGTCATGCTTTTTTACACCTTTTACACCTTTCTCCACATTTCTCAAGTCACCA	1298		
Db	121	ACTTTGAGTGGTCATGCTTTTTTACACCTTTTACACCTTTCTCCACATTTCTCAAGTCACCA	180		
QY	1299	AGTCATAGCCATGAAGAACCCAGCAATGGAAATGAAAGAGGACCACTTTTCAGTCATCTG	1358		
Db	181	AGTCATAGCCATGAAGAACCCAGCAATGGAAATGAAAGAGGACCACTTTTCAGTCATCTG	240		
QY	1359	TCCTCTCAAAACATAGAAGTAGTGCCTATTTTGATTCACGTGGAAGGGTCTTAACAGCT	1418		
Db	241	TCCTCTCAAAACATAGAAGTAGTGCCTATTTTGATTCACGTGGAAGGGTCTTAACAGCT	300		
QY	1419	CTAGGAGGCGCTGATTTTCATGTTTCTGTGTGAACATGTCTCACAATTCATCAAAACAATTT	1478		
Db	301	CTAGGAGGCGCTGATTTTCATGTTTCTGTGTGAACATGTCTCACAATTCATCAAAACAATTT	360		
QY	1479	AAAGATATAGAAGAAAAGAAATCAGAAAGAACCTGAAAATGATGATGTGGAGATTAAAG	1538		
Db	361	AAAGATATAGAAGAAAAGAAATCAGAAAGAACCTGAAAATGATGATGTGGAGATTAAAG	420		
QY	1539	AAGCAGTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATGAGGAGAGTAGATACAGAT	1598		
Db	421	AAGCAGTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATGAGGAGAGTAGATACAGAT	480		
QY	1599	GATCCAACCTGAAGGCTATTTACGAGCAGACATCAAGAGCCCTCCCACCTTTGATTTCTCAG	1658		
Db	481	GATCCAACCTGAAGGCTATTTACGAGCAGACATCAAGAGCCCTCCCACCTTTGATTTCTCAG	540		
QY	1659	CAGCCTCGAGTCTTGGAAAGAAAGAGGTCAATGATAGCTCATGCTCATCCACAGGAAGTC	1718		
Db	541	CAGCCTCGAGTCTTGGAAAGAAAGAGGTCAATGATAGCTCATGCTCATCCACAGGAAGTC	600		
QY	1719	TACAAATGAATATGTACCCAGAGGTGCAAGATAAATGCCATTTCCACGATACA	1778		
Db	601	TACAAATGAATATGTACCCAGAGGTGCAAGATAAATGCCATTTCCACGATACA	660		
QY	1779	CTCGCCAGTCAGACGATCTCATTCACACCAATCATGACTACCATTTCCACGATACA	1838		
Db	661	CTCGCCAGTCAGACGATCTCATTCACACCAATCATGACTACCATTTCCACGATACA	720		
QY	1839	CACCACCAACCAACCAATCTTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGCTG	1898		
Db	721	CACCACCAACCAACCAATCTTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGCTG	780		
QY	1899	AAAGATCCCGCGGTGCGCCACTTTGGCGCTGGATGGTGATATGGGTGATGGCCTGCACAAAT	1958		
Db	781	AAAGATCCCGCGGTGCGCCACTTTGGCGCTGGATGGTGATATGGGTGATGGCCTGCACAAAT	840		

```
OY 1959 TTCACGGATGGCCTAGCAATTTGGTCTGCTTTTACTGAAGCTTATCAAGTGGTTAAAGT 2018
DB 841 TTCACGGATGGCCTAGCAATTTGGTCTGCTTTTACTGAAGCTTATCAAGTGGTTAAAGT 900
OY 2019 ACTTCTGTTGCTGCTTCTCTCATGATGCTGCTCATGAATAGTGACTTTGCTGTCTTA 2078
DB 901 ACTTCTGTTGCTGCTTCTCTCATGATGCTGCTCATGAATAGTGACTTTGCTGTCTTA 960
OY 2079 CTAAAGCCTGGCATCAGCCTTTAAGCAGGCTGCTCTTTAATGATGCTGCTGCTGCTG 2138
DB 961 CTAAAGCCTGGCATCAGCCTTTAAGCAGGCTGCTCTTTAATGATGCTGCTGCTGCTG 1020
OY 2139 GCGTATCTTGAATGGCAACAGGAATTTTCATTGGTCATTATGCTGAAATGTTCTATG 2198
DB 1021 GCGTATCTTGAATGGCAACAGGAATTTTCATTGGTCATTATGCTGAAATGTTCTATG 1080
OY 2199 TGGATATTTGCATCTACTGCTGCTTATTCATGATGCTGCTGCTGCTGCTGCTGCTG 2258
DB 1081 TGGATATTTGCATCTACTGCTGCTTATTCATGATGCTGCTGCTGCTGCTGCTGCTG 1140
OY 2259 GAAATGCTGCACATGATGCTAGTACCATGATGATGATGATGATGATGATGATGATG 2318
DB 1141 GAAATGCTGCACATGATGCTAGTACCATGATGATGATGATGATGATGATGATGATG 1200
OY 2319 CAGATGCTGGGATGCTTTGGGCTTTGGGATTTGGAATTTGCTGCTGCTGCTGCTGCT 1259
DB 1201 CAGATGCTGGGATGCTTTGGGCTTTGGAATTTGCTGCTGCTGCTGCTGCTGCTGCT 2378
OY 2379 AAAATGCTGTTTCTGATAAATTTCTAGTTAAAGTTTAAATGCTAGATGAGCT 2430
DB 1260 AAAATGCTG-TTCGTATAATTTCTAGTTAAAGTTTAAATGCTAGATGAGCT 1310

RESULT 10
AC091060/c
LOCUS
DEFINITION Homo sapiens chromosome 18, clone RP11-723J4, complete sequence.
ACCESSION AC091060
VERSION AC091060.10 GI:16646826
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
TITLE Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M.,
JOURNAL Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K.,
Diaz, J.S., Dodge, S., Faro, S., Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Lamazzes, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Rett, R., Rosetti, M., Riley, R., Rise, C., Rogov, P.,
Seaman, S., Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Topham, K., Travers, M., Talamas, J., Tesfaye, S., Theodore, J.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (28-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 4, 2001 this sequence version replaced gi:16118168.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L12269
Center clone name: 723-J_4
-----
Location/Qualifiers
1. 221941
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-723J4"
/clone_lib="RPC1-11 Human Male BAC"
240. 261
/notes="<30 qual SNGL region"
322. 326
/notes="<30 qual SNGL region"
635. 670
/repeat_family="(TTTA)n"
complement(674. 963)
/repeat_family="AluSg"
688. 715
/notes="<30 qual SNGL region"
951. 955
/notes="<30 qual SNGL region"
1010. 1013
/notes="<30 qual SNGL region"
1124. 1441
/repeat_family="AluSx"
2833. 2926
/repeat_family="(CA)n"
complement(2960. 3257)
/repeat_family="AluJo"
complement(3291. 3371)
/repeat_family="L2"
3807. 4209
/repeat_family="L1MD1"
complement(4298. 4597)
/repeat_family="AluSx"
```

```
repeat_region 5434..5597
repeat_region /rpt_family="FRAM"
repeat_region complement(6415..6580)
repeat_region /rpt_family="MIR"
repeat_region complement(7324..7451)
repeat_region /rpt_family="MLT1H"
repeat_region 7505..7750
repeat_region /rpt_family="AluJb"
repeat_region complement(7768..7984)
repeat_region /rpt_family="MLT1H"
repeat_region 8022..8284
repeat_region /rpt_family="MLT1J"
repeat_region complement(8865..9153)
repeat_region /rpt_family="AluSc"
repeat_region 9584..9680
repeat_region /rpt_family="AluSg/x"
repeat_region 11358..11524
repeat_region /rpt_family="AluSx"
repeat_region 11555..11586
repeat_region /rpt_family="(TAAA)n"
repeat_region 11587..11750
repeat_region /rpt_family="L1ME2"
repeat_region 11838..12141
repeat_region /rpt_family="AluSx"
repeat_region complement(12195..12762)
repeat_region /rpt_family="MLT1E2"
repeat_region 12824..12950
repeat_region /rpt_family="AluSg/x"
repeat_region 12955..13254
repeat_region /rpt_family="AluSx"
repeat_region 13256..13920
repeat_region /rpt_family="L1MB2"
repeat_region 13921..14181
repeat_region /rpt_family="AluSx"
repeat_region 14182..14516
repeat_region /rpt_family="L1MB2"
repeat_region 14572..14886
repeat_region /rpt_family="L1ME2"
repeat_region 14905..14936
repeat_region /rpt_family="MER58B"
repeat_region 14937..15226
repeat_region /rpt_family="AluSg"
repeat_region 15227..15376
repeat_region /rpt_family="MER58B"
repeat_region 15377..16032
repeat_region /rpt_family="L1ME2"
repeat_region complement(16039..16172)
repeat_region /rpt_family="L1MCA"
repeat_region complement(16197..17024)
repeat_region /rpt_family="L1MCA"
repeat_region complement(17025..17305)
repeat_region /rpt_family="AluSx"
repeat_region complement(17306..17710)
repeat_region /rpt_family="L1MCA"
repeat_region 17710..18046
repeat_region /rpt_family="L1ME2"
repeat_region 18047..18330
repeat_region /rpt_family="GA-rich"
repeat_region 18336..19013
repeat_region /rpt_family="L1ME2"
repeat_region 19014..19286
repeat_region /rpt_family="AluYa5"
repeat_region 19287..19895
repeat_region /rpt_family="L1ME2"
repeat_region 19896..20066
repeat_region /rpt_family="AluJb"
repeat_region 20070..20562
repeat_region /rpt_family="L1ME2"
repeat_region 20680..20722
repeat_region /rpt_family="L1ME2"
repeat_region 20817..20961
repeat_region /rpt_family="L1MC5"
repeat_region complement(20990..21142)
```

```
Query Match 32.9% Score 1137.8; DB 9; Length 221941;
Best Local Similarity 98.9%; Pred. No. 3.5e-245;
Matches 1209; Conservative 0; Mismatches 7; Indels 7; Gaps 6;

QY 2245 TTGATATGGTACCCTGAATGCTGCACAATGATGCTAGTGACCATGATGATACCGCTGGG 2304
Db 170112 TTTTAAGGTACTGAATGCTGCACAATGATGCTAGTGACCATGATGATACCGCTGGG 170053
QY 2305 GGTATTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGAATTATGTTACTATT 2364
Db 170052 GGTATTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGAATTATGTTACTATT 169993
QY 2365 CCATATTTGAACATAAATTCGTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGA 2424
Db 169992 CCATATTTGAACATAAATTCGTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGA 169933
QY 2425 GTAGCTTAAAAAGTTGTCATAGTTTTCAGTAGTTCATAGGAGATGAGTTGATGCTGTA 2484
Db 169932 GTAGCTTAAAAAGTTGTCATAGTTTTCAGTAGTTCATAGGAGATGAGTTGATGCTGTA 169873
QY 2485 CTATGCAGGTTTAAAGTTAGTGGGTTTGTGATTTTGTATTTGATTAATGCTGCTGTT 2544
Db 169872 CTATGCAGGTTTAAAGTTAGTGGGTTTGTGATTTTGTATTTGATTAATGCTGCTGTT 169813
QY 2545 ACAAAGTCAGTTAAAGTACGTTTAAATATTTAAAGTTATTTCTATCTTGGAGATAAATCT 2604
Db 169812 ACAAAGTCAGTTAAAGTACGTTTAAATATTTAAAGTTATTTCTATCTTGGAGATAAATCT 169753
QY 2605 STATGTGCAATTCACCGGTATTTACAGTTTATATGTAACAAGAGATTTGGCATGACAT 2664
Db 169752 STATGTGCAATTCACCGGTATTTACAGTTTATATGTAACAAGAGATTTGGCATGACAT 169693
QY 2665 GTTCTGTATGTTTCAGGGAAAAATGCTTTAATGCTTTTTCAGAACTAACACAGTTATT 2724
Db 169692 GTTCTGTATGTTTCAGGGAAAAATGCTTTAATGCTTTTTCAGAACTAACACAGTTATT 169633
QY 2725 CCTATCTGGATTTTAGTCTCTGAAGAACTGCTGGTGTGTTTAGGAATAAGAAATGTCATG 2784
Db 169632 CCTATCTGGATTTTAGTCTCTGAAGAACTGCTGGTGTGTTTAGGAATAAGAAATGTCATG 169573
QY 2785 AAGCCTTAAATACCAACAAGCTTATCTGAATTTAAGCAAGAAATAAAGGAGAAAAGA 2844
Db 169572 AAGCCTTAAATACCAACAAGCTTATCTGAATTTAAGCAAGAAATAAAGGAGAAAAGA 169513
QY 2845 GAGAATCTCAGAAATTTGGGGAGGCATAGATTCTTATAAAAATCACAANAATTTGTTGTA 2904
Db 169512 GAGAATCTCAGAAATTTGGGGAGGCATAGATTCTTATAAAAATCACAANAATTTGTTGTA 169453
QY 2905 TTAGAGGGGAGAAATTTAGAAATTAAGTATATAAAGGCAGAAATTAGTATAGATACATCA 2964
Db 169452 TTAGAGGGGAGAAATTTAGAAATTAAGTATATAAAGGCAGAAATTAGTATAGATACATCA 169393
QY 2965 TTAACATTTTGTCCAGGATTTTCCCGTAAAAAGCTAGTGAGCAC-TCTCATATACTA 3023
Db 169392 TTAACATTTTGTCCAGGATTTTCCCGTAAAAAGCTAGTGAGCAC-TCTCATATACTA 169333
```


QY	3024	ATT--AGTGTACATTTAACTTTGTGTATAATACAGAAATCTAAATATATTTAATGAATCAA	3081
Db	169332	ATTTAGTTGTACATTTAACTTTGTGTATAATACAGAAATCTAAATATATTTAATGAATCAA	169273
QY	3082	GCAATATA-CACTTGACCAAGAAATTTGGAAATTTCAAATGTCCTCGCGGTATATACCA	3140
Db	169272	GCAATATATCAGTTGACCAAGAAATTTGGAAATTTCAAATGTCCTCGCGGTATATACCA	169214
QY	3141	GATCAGTACAGTCAGTAG-TTTATGATATCACCAGACTGGGTTATTGCCAAGTTATATATC	3199
Db	169213	GATCAGTACAGTCAGTAGTTTATGTTATCACCAGACTGGGTTATTGCCAAGTTATATATC	169154
QY	3200	ACCAAAGCTGTATGACTCGATGTTCTGGTTACCTGGTTTACAAAATATCAGAGTAGTA	3259
Db	169153	ACCAAAGCTGTATGACTCGATGTTCTGGTTACCTGGTTTACAAAATATCAGAGTAGTA	169094
QY	3260	AACTTTGATATATATGAGGATATTTAAACATACACTAGTATCATTTGATTCGATTCAGA	3319
Db	169093	AACTTTGATATATATGAGGATATTTAAACATACACTAGTATCATTTGATTCGATTCAGA	169034
QY	3320	AAGTACTTTGATATCTCTCAGTGCTTCAGTCTCATCTGTTGAGCAATTTGTC-TTTATAT	3378
Db	169033	AAGTACTTTGATATCTCTCAGTGCTTCAGTCTCATCTGTTGAGCAATTTGTC-TTTATAT	168974
QY	3379	ACGGTACTGTAGCCATACCTAGGCGCTGCTGTGGCAATCTCTAGATGTTCTTTTTACAC	3438
Db	168973	ACGGTACTGTAGCCATACCTAGGCGCTGCTGTGGCAATCTCTAGATGTTCTTTTTACAC	168914
QY	3439	AATAAATTCCTTATATCAGCTTG	3461
Db	168913	AATAAATTCCTTATATCAGCTTG	168891

RESULT 11	ACCESSION	ORGANISM
AP001905	VERSION	
LOCUS	KEYWORDS	
DEFINITION	SOURCE	

AP001905 224788 bp DNA linear HTG 13-JUL-2000
Homo sapiens chromosome 18 clone RP11-723J4
DRAFT SEQUENCE, 23 unordered pieces.
AP001906
AP001907 3 GI:9188521
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens DNA, clone:RP11-723J4.
Homo

REFERENCE AUTHORS	TITLE JOURNAL REFERENCE AUTHORS	TITLE JOURNAL.
----------------------	--	-------------------

Mammalia: Eutheria: Primates: Catarrhini; Hominoideae: Homo.
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 1 (bases 1 to 224788) Toyoki, Y., Watanabe, H. and Sakaki, Y.
 2 (bases 1 to 224788)

Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail: hattori@gscc.riken.go.jp, URL: <http://hgpc.gsc.riken.go.jp/>,
 Tel: 81-42-778-9933, Fax: 81-42-778-9924)
 On Jul 14, 2000 this sequence version replaced gi:8117556.
 ***** Genomic Coordinates *****

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: <http://hgsc.gsc.riken.go.jp/>
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: Rp11-723J4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-Terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 216741 bases at least Q40

Consensus quality: 219708 bases at least Q30
 Consensus quality: 221297 bases at least Q20
 Insert size: 222588; sum-of-contigs
 Quality coverage: 9.33x in Q20 bases; sum-of-

NOTE: This is a 'working draft' sequence. It currently contains 23 contigs. The true order of the pieces is not known and the order in this sequence record is arbitrary. Gaps between contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

42545	42444	contig of	42444	bp	in length
64760	84659	contig of	21115	bp	in length
82977	84876	contig of	18117	bp	in length
101007	109006	contig of	17930	bp	in length
118052	117951	contig of	16945	bp	in length
131604	131503	contig of	13452	bp	in length
141638	141537	contig of	9934	bp	in length
152990	152889	contig of	11252	bp	in length
161919	161818	contig of	8029	bp	in length
171084	170983	contig of	9065	bp	in length
187340	187239	contig of	7156	bp	in length
187673	187572	contig of	9233	bp	in length
193133	193032	contig of	5360	bp	in length
198943	198842	contig of	5710	bp	in length
203133	203032	contig of	4051	bp	in length
207884	207783	contig of	4690	bp	in length
211529	211428	contig of	3545	bp	in length
215692	215591	contig of	4063	bp	in length
217933	217832	contig of	2141	bp	in length
220611	220510	contig of	2578	bp	in length
222136	222035	contig of	1425	bp	in length
223756	223655	contig of	1320	bp	in length
	224788	contig of	1033	bp	in length

NOTE: This is a working draft.

* NOTE: This is a 'working draft' of 1033 bp in length consisting of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N' and the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 24444: contig of 42444 bp in length
 * 42545: gap of 100 bp
 * 64659: contig of 22115 bp in length
 * 64660: gap of 100 bp
 * 82876: contig of 18117 bp in length
 * 82877: gap of 100 bp
 * 100907: contig of 17930 bp in length
 * 101007: gap of 100 bp
 * 117951: contig of 16945 bp in length
 * 117952: gap of 100 bp
 * 118051: gap of 100 bp
 * 131503: contig of 13452 bp in length
 * 131504: gap of 100 bp
 * 131604: gap of 100 bp
 * 141537: contig of 9934 bp in length
 * 141538: gap of 100 bp
 * 141638: contig of 11252 bp in length
 * 152890: gap of 100 bp
 * 152899: contig of 8829 bp in length
 * 152990: gap of 100 bp
 * 161818: contig of 8829 bp in length
 * 161918: gap of 100 bp
 * 161919: gap of 100 bp
 * 170983: contig of 9065 bp in length
 * 170984: gap of 100 bp
 * 171084: contig of 7156 bp in length
 * 178240: gap of 100 bp
 * 178339: contig of 9233 bp in length
 * 178340: gap of 100 bp
 * 187572: contig of 5360 bp in length
 * 187573: gap of 100 bp
 * 193032: contig of 5360 bp in length
 * 193033: gap of 100 bp
 * 193132: contig of 5710 bp in length
 * 193133: gap of 100 bp
 * 198842: contig of 5710 bp in length
 * 198843: gap of 100 bp
 * 203032: contig of 4090 bp in length
 * 203033: gap of 100 bp


```

OY 3320 AGTACTTTGATATCTCTCAGTGGCTTCACTGCTATCATGTGAGCAATGTC-TTTATAT 3378
|||||
Db 20682 AAGTACTTTGATATCTCTCAGTGGCTTCACTGCTATCATGTGAGCAATGTC-TTTATAT 20741
|||||
OY 3379 ACGGTACTCTAGCCATAGCCCTGCTGTGGCAATCTCTAGATGTTCTTTTACAC 3438
|||||
Db 20742 ACGGTACTCTAGCCATAGCCCTGCTGTGGCAATCTCTAGATGTTCTTTTACAC 20801
|||||
OY 3439 AATAAATCTTATATCAGCTTG 3461
|||||
Db 20802 AATAAATCTTATATCAGCTTG 20824
|||||

RESULT 12
AP001158
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-701C9 map 18q12, WORKING
DRAFT SEQUENCE, 15 unordered pieces.
ACCESSION AP001158
VERSION AP001158.3 GI:9188478
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE Homo sapiens DNA, clone: RP11-701C9.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Vada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 160,170 genomic DNA of 18q12
Published Only in Database (2000)
2 (bases 1 to 160170)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Vada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (10-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8535, Japan
(E-mail: hattori@psc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-42-778-9923, Fax: 81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8118726.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@psc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-701C9
----- Summary Statistics
Sequencing Vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 155820 bases at least Q40
Consensus quality: 157459 bases at least Q30
Consensus quality: 158091 bases at least Q20
Insert size: 158770; sum-of-contigs
Quality coverage: 10.41x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
15 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 29590 contig of 29590 bp in length
29691 49906 contig of 20216 bp in length
50007 69120 contig of 19114 bp in length
69221 84192 contig of 14972 bp in length
84293 99890 contig of 15598 bp in length
99991 110467 contig of 10477 bp in length
110568 118491 contig of 7924 bp in length

NOTE: This is a 'working draft' sequence. It currently
consists of 15 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 29590: contig of 29590 bp in length
29591 29690: gap of 100 bp
29691 49906: contig of 20216 bp in length
49907 50006: gap of 100 bp
50007 69120: contig of 19114 bp in length
69121 69220: gap of 100 bp
69221 84192: contig of 14972 bp in length
84193 84292: gap of 100 bp
84293 99890: contig of 15598 bp in length
99891 99990: gap of 100 bp
99991 110467: contig of 10477 bp in length
110468 110567: gap of 100 bp
110568 118491: contig of 7924 bp in length
118492 118591: gap of 100 bp
118592 126449: contig of 7858 bp in length
126450 126549: gap of 100 bp
126550 134958: contig of 8409 bp in length
134959 135058: gap of 100 bp
135059 141053: contig of 5995 bp in length
141054 141153: gap of 100 bp
141154 145920: contig of 4767 bp in length
145921 146020: gap of 100 bp
146021 149439: contig of 3419 bp in length
149440 149539: gap of 100 bp
149540 154173: contig of 4634 bp in length
154174 154273: gap of 100 bp
154274 157616: contig of 3343 bp in length
157617 157716: gap of 100 bp
157717 160170: contig of 2454 bp in length.

FEATURES
Source
1..160170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="RP11-701C9"

misc_feature
1..29590
/note="assembly_fragment"
29691..49906
/note="assembly_fragment clone_end:T7 vector_side:right"
50007..69120
/note="assembly_fragment"
69221..84192
/note="assembly_fragment"
84293..99890
/note="assembly_fragment"
99991..110467
/note="assembly_fragment"
110568..118491
/note="assembly_fragment clone_end:sp6 vector_side:right"
118592..126449
/note="assembly_fragment"
126550..134958
/note="assembly_fragment"
135059..141053
/note="assembly_fragment"
141154..145920
/note="assembly_fragment"

```

misc_feature	/note="assembly_fragment"
	146021 .149439
misc_feature	/note="assembly_fragment"
	149540. .154173
misc_feature	/note="assembly_fragment"
	154274. .157616
misc_feature	/note="assembly_fragment"
	157717. .160170
	/note="assembly_fragment"
BASE COUNT	46219 a 33232 c 33436 g 45883 t
ORIGIN	1400 others

Query Match	23.2%	Score 804;	DB 2;	Length 160170;
Best Local Similarity	87.6%;	Pred. No. 4.7e-170;		
Matches 836;	Conservative 0;	Mismatches 115;	Indels 3;	Gaps 2;
QY	2245	TTGATATGGTACTCGAATGCTGCACAATGATGCTAGTGACCATGGATGGATGACCGCTGGG	2304	
DB	49069			
QY	2305	GGTATTTCTTTTACAGAATGCTGGGATGCTTTTGGGTTTTGGAAATATGTTACTTATTT	2364	
DB	49129	GGTATTTCTTTTACAGAATGCTGGGATGCTTTTGGGTTTTGGAAATATGTTACTTATTT	49188	
QY	2365	CCATATTTCAACATAAAAATCGTTTCTGTATAAATTTCTAGTTAAGGTTTAAATGCTAGA	2424	
DB	49189	CCATATTTGAACATAAAATCGTTTCTGTATAAATTTCTAGTTAAGGTTTAAATGCTAGA	49248	
QY	2425	GTAGCTTAAAAAGTTGTCATAGTTTCACTAGGTCATAGGGAGATGAGTTTGTATGCTGTA	2484	
DB	49249	GTAGCTTAAAAAGTTGTCATAGTTTCACTAGGTCATAGGGAGATGAGTTTGTATGCTGTA	49308	
QY	2485	CTATGCAGCGTTTAAAGTTACTGGGTTTTGNGANTTTTGTAATTCGAATATGCTGTCTGTT	2544	
DB	49309	CTATGCAGCGTTTAAAGTTACTGGGTTTTGNGANTTTTGTAATTCGAATATGCTGTCTGTT	49368	
QY	2545	ACAAAGTCAGTTAAAGGTACGTTTTTAATTAATTTAAAGTTATTTCTATCTTGGAGATAAAATCT	2604	
DB	49369	ACAAAGTCAGTTAAAGGTACGTTTTTAATTAATTTAAAGTTATTTCTATCTTGGAGATAAAATCT	49428	
QY	2605	GTATGTGCAATTCACCGTATTACCAAGTTATTATGTATGAACAAGAGATTTGGCATGACAT	2664	
DB	49429	GTATGTGCAATTCACCGTATTACCAAGTTATTATGTATGAACAAGAGATTTGGCATGACAT	49488	
QY	2665	GTTCTGTATGTTTCAGGGAAAAATGTCTTTAAATGCTTTTTCAAGAACTAAACACAGTTATT	2724	
DB	49489	GTTCTGTATGTTTCAGGGAAAAATGTCTTTAAATGCTTTTTCAAGAACTAAACACAGTTATT	49548	
QY	2725	CCATACTGGATTTTACGCTCTCAAGAACTGCTGCTGTTTTAGAAATAAGAAATGTGCATG	2784	
DB	49549	CCATACTGGATTTTACGCTCTCAAGAACTGCTGCTGTTTTAGAAATAAGAAATGTGCATG	49608	
QY	2785	AAGCCTAAAATACCAAGAAAGCTTTATACTGAATTTTAGCAAAAGAAATAAAGAGAGAAAGA	2844	
DB	49609	AAGCCTAAAATACCAAGAAAGCTTTATACTGAATTTTAGCAAAAGAAATAAAGAGAGAAAGA	49668	
QY	2845	GAGAATCTCAGAAATGGGAGGCATAGATTCCTTATAAAATCACAAAATTTGCTGTGAAA	2904	
DB	49669	GAGAATCTCAGAAATGGGAGGCATAGATTCCTTATAAAATCACAAAATTTGCTGTGAAA	49728	
QY	2905	TTAGAGGGGAGAAATTTAGAAATTAAGTATATAAAAGGCAGAAATTTAGTATAGAGTACATTC	2964	
DB	49729	TTAGAGGGGAGAAATTTAGAAATTAAGTATATAAAAGGCAGAAATTTAGTATAGAGTACATTC	49788	
QY	2965	TTAAACATTTTTCGACGATTATTTCCCGTAAAAAGCTAGTGAGCAC-TCTCATATACCTA	3023	
DB	49789	TTAAACATTTTTCGACGATTATTTCCCGTAAAAAGCTAGTGAGCAC-TCTCATATACCTA	49848	
QY	3024	ATT--AGTGTACATTTAACTTTGTATAATACAGAAATCTAAATATATTTAAATGAATTC	3081	
DB	49849	ATTAGTGTACATTTAACTTTGTATAATACAGAAATCTAAATATATTTAAATGAATTC	49908	
QY	3082	GCAATATACACTTGACCAGAAATTTGGAATTTCAAAATGTTCCGTGGGTTTATATACCA	3141	

[illegible]

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshcarl,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 193437)
Worley,K.C.
Direct Submission
Submitted (08-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 193437)
Worley,K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:21703549.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBBL
Center clone name: CH230-5P17
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 143517 bases at least Q40
Consensus quality: 150307 bases at least Q30
Consensus quality: 154257 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1172: contig of 1172 bp in length
* 1173 1272: gap of unknown length
* 1273 2487: contig of 1215 bp in length
* 2488 2587: gap of unknown length
* 2588 4394: contig of 1807 bp in length
* 4395 4494: gap of unknown length
* 4495 5815: contig of 1321 bp in length
* 5816 5915: gap of unknown length
* 5916 7626: contig of 1711 bp in length
* 7627 7726: gap of unknown length
* 7727 9206: contig of 1480 bp in length
* 9207 9307: gap of unknown length
* 9307 10449: contig of 1143 bp in length
* 10450 10549: gap of unknown length
* 10550 11609: contig of 1060 bp in length
* 11610 11709: gap of unknown length
* 11710 12957: contig of 1248 bp in length
* 12958 13057: gap of unknown length
* 13058 14501: contig of 1444 bp in length
* 14502 14601: gap of unknown length
* 14602 15762: contig of 1161 bp in length
* 15763 15862: gap of unknown length
* 15863 16877: contig of 1015 bp in length
* 16878 16977: gap of unknown length
* 16978 18761: contig of 1784 bp in length
* 18762 18861: gap of unknown length
* 18862 20537: contig of 1676 bp in length
* 20538 22730: contig of 2093 bp in length
* 22731 22830: gap of unknown length
* 22831 25262: contig of 2432 bp in length
* 25263 25362: gap of unknown length
* 25363 27262: contig of 1900 bp in length
* 27263 27362: gap of unknown length
* 27363 28524: contig of 1162 bp in length
* 28525 28624: gap of unknown length
* 28625 30233: contig of 1609 bp in length
* 30234 30334: gap of unknown length
* 30334 31884: contig of 1550 bp in length
* 31884 31984: gap of unknown length
* 31984 33437: contig of 1453 bp in length
* 33437 33537: gap of unknown length
* 33537 34740: contig of 1203 bp in length
* 34740 34839: gap of unknown length
* 34840 36918: contig of 2078 bp in length
* 36918 37018: gap of unknown length
* 37018 38174: contig of 1156 bp in length
* 38174 38274: contig of 1679 bp in length
* 38274 39953: gap of unknown length
* 39953 40052: gap of unknown length
* 40052 42001: contig of 1948 bp in length
* 42001 42100: gap of unknown length
* 42100 43702: contig of 1602 bp in length
* 43702 43802: gap of unknown length
* 43802 46149: contig of 2347 bp in length
* 46149 46249: gap of unknown length
* 46249 47598: contig of 1349 bp in length
* 47598 47698: gap of unknown length
* 47698 49211: contig of 1512 bp in length
* 49211 49310: gap of unknown length
* 49310 50780: contig of 1470 bp in length
* 50780 50880: gap of unknown length
* 50880 50881: contig of 1970 bp in length
* 50881 52850: gap of unknown length
* 52850 52950: gap of unknown length
* 52950 53981: contig of 1031 bp in length
* 53981 54081: gap of unknown length
* 54081 55821: contig of 1740 bp in length
* 55821 55921: gap of unknown length
* 55921 57909: contig of 1988 bp in length
* 57909 58009: gap of unknown length
* 58009 60388: contig of 2379 bp in length
* 60388 60488: gap of unknown length
* 60488 62460: contig of 1972 bp in length
* 62460 62560: gap of unknown length
* 62560 64500: contig of 1940 bp in length
* 64500 64600: gap of unknown length
* 64600 66889: contig of 2289 bp in length
* 66889 69589: gap of unknown length
* 69589 69689: gap of unknown length
* 69689 71653: contig of 1964 bp in length
* 71653 71753: gap of unknown length
* 71753 75273: contig of 3520 bp in length
* 75273 75373: gap of unknown length
* 75373 77772: contig of 2399 bp in length
* 77772 77872: gap of unknown length
* 77872 81413: contig of 3541 bp in length
* 81413 81513: gap of unknown length
* 81513 85249: contig of 3736 bp in length
* 85249 85349: gap of unknown length
* 85349 87366: contig of 2017 bp in length
* 87366 87466: gap of unknown length
* 87466 90125: contig of 2659 bp in length
* 90125 87467

Db 108242 ATTCCAGAACCTCAAAATGACAAATATCCAGGAG 108273

Search completed: November 8, 2002, 19:49:16
Job time : 7252.03 secs

Query Match		12.4%;	Score 429.2;	DB 2;	Length 193437;
Best Local Similarity		75.7%;	Pred. No. 1.1e-85;		
Matches 615;		Conservative 0;	Mismatches 128;	Indels 69;	Gaps 4;
QY 127	ACGAAGCGCAATGGCGAGGAAGTATCTGTATCTTGATCTCTGACCTTTGGCCCTCTCTG 186				
Db 107519	AGGAAGGTGCAATGGCCACAAATTTATCTGTATCATGATCTTGACCTTTGCCCTTTGGG 107578				
QY 187	TCACAAATCCCTTCATGAACATAA---AGCAGCTGCTTTCCCCAGACCACTGAGAAAA 243				
Db 107579	TTACAAATCCCTTCATGAACATAAATCAACAGCTGCTTTCTCTCAGACCACTGAGAAAA 107638				
QY 244	TTAGTCGGAATGGGAATCTGGCATTAAATGTTGACTTTGGCAATTTCCACACGGCAATATC 303				
Db 107639	TTAATTCAAATTCGGGAATCTGGCAATTAATGTTGACCTGGCAGTTACCATGCGAGCACATC 107698				
QY 304	ATCTACACAGCTTTTCTACCGCTATGGAGAAAAATTAATTTTGTTCAGTTGAAGGGTTCA 363				
Db 107699	ATCTGCAACAGCTTTTCTACCGCTATGGAGAGAAATGACTCTCTTGTTCAGTTGAAGGGTTCA 107758				
QY 364	GAATTTACTTTCAAAATATAGGCATAGATTAAGATTAAAGAAATCCATATACACCATGACC 423				
Db 107759	GAATTTACTTTCAAAATATAGGCATAGATTAAGATTAAAGAAATCCATATACACCATGACC 107818				
QY 424	ACGACCATCACTCAGACCCAGGAGCATCACTCAGACCATGAGGCTCACTCAGACCATGAGC 483				
Db 107819	AC-----GAGC 107824				
QY 484	ATCACTCAGACCAAGCAGCTCACTTGACCATGATCATCACTCTCACCATAATCATGCTG 543				
Db 107825	GCCACTCTGACCAGAGCGGCACTCTGACCAGCGCCACTCCCCACCGAGGTCATGCTG 107884				
QY 544	CTTCTGGTAAAAATAAGCGAAAGCTCTTTGCCAGACCACTGACTCAGATAGTTTCAGGTA 603				
Db 107885	CCGCTGGGAAAAACAGTCGGAAAGCCCTTTTGCCAGACCTTGACTCTGATAATTCAGGTA 107944				
QY 604	AAGATCTAGAAACAGCGAGGGAAGGAGCTCACCAGCAGCAACATGCCAGTGGTAGAA 663				
Db 107945	AAATCGGAACACCGCCAGGTAAAGGATCTCGCCACGAGAGCATGTGAATGGCAGGA 108004				
QY 664	GGATGTCAAGGACAGTGTTCAGTGTAGTCAAGTACCTCAACTGTGTACAACACTGTCT 723				
Db 108005	GGAATGGCAAGGAGAGTGAAGCTCTAGTGAAGTACCTCGGCAGTTTACAACACTGTCT 108064				
QY 724	CTGAAGGAACCTCACTTTCTAGACACAATAGAGACTCCAGACCTGGAAAACTCTTCCCCA 783				
Db 108065	CTGAAGGAACCTCACTTTCTAGACACAATAGAGACTCCAAAAACCCGGGAGAC---GCACCA 108121				
QY 784	AAGATGTAGCAGCTCCACTCCACCCAGTGTACATCAAGAGCGCGGTGAGCCGGCTG- 842				
Db 108122	AAGATATAAACCTTCTACCCCAACCCAGCATCACAGAGAGCCGAGTGGCGCGCTGA 108181				
QY 843	-----GCTGTAGGAAACAAATGAATCTGTGAGTGAGCCGCCAAAAGGCTTTTATGT 894				
Db 108182	GCGGACTAGCTAGGAGAAAGGCAACGATTCGTGTGAGTGAGCCGAGGAGGCTTCATGT 108241				
QY 895	ATTCCAGAACACAAATGAAAATCTTCAGGAG 926				

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2002, 17:37:02 ; Search time 3433.14 Seconds
(without alignments)
16326.890 Million cell updates/sec

Title: US-09-642-034-1
Perfect score: 3461
Sequence: 1 ctctgtccgaattcgccagc.....aaattccttatcagcttg 3461

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_nam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	877.6	25.4	1074	13	BM480018
2	800.6	23.1	915	14	BQ716456
3	799.6	23.1	872	9	AU120027
4	738.8	21.3	1059	13	BM557996
5	724	20.9	810	13	BM456317
6	711	20.5	712	14	BM721841

7	694	20.1	782	12	BG571701
8	692	20.0	701	14	BQ027619
9	676.8	19.6	728	12	BG403446
10	661.6	19.1	830	13	B1222452
11	652.6	18.9	697	9	AL039253
12	652	18.8	1135	13	BM560789
13	649.8	18.8	674	14	BM978669
14	633.2	18.3	673	9	A1744245
15	627.6	18.1	957	12	BG168169
16	625.2	18.1	720	14	BQ013982
17	617.2	17.8	927	14	BQ926045
18	613.4	17.7	791	12	BG547794
19	605.4	17.5	746	12	BG530601
20	597.6	17.3	604	14	BQ581865
21	591.8	17.1	857	12	BF032013
22	591.4	17.1	743	9	A1907176
23	586.8	17.0	766	14	BM948026
24	576	16.6	925	12	BF983458
25	571.4	16.5	581	9	A1635070
26	567.4	16.4	829	9	AU079709
27	566.8	16.4	751	14	BM963737
28	566.2	16.4	887	13	B1150324
29	564	16.3	781	14	BM949686
30	561.4	16.2	684	14	BQ009520
31	560	16.2	670	14	BM999466
32	559	16.2	559	10	AW771378
33	555.2	16.0	708	12	BF978788
34	551.8	15.9	765	12	BG168944
35	538.6	15.6	948	14	BQ955836
36	538.4	15.6	733	10	BE391889
37	526.6	15.2	580	12	BF915495
38	523.6	15.1	836	12	BF790668
39	517.8	15.0	849	12	BF680209
40	510.4	14.7	797	12	BF679305
41	505.2	14.6	682	14	BM949679
42	505.2	14.6	834	12	BF669257
43	503.4	14.5	950	14	BQ433766
44	502.2	14.5	518	10	AW178411
45	501.8	14.5	513	10	AW178409

ALIGNMENTS

RESULT 1
BM480018
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

BM480018
AGENCOURT_6424145 NIH_MGC_67 Homo sapiens cdna clone IMAGE:5491572
5', mRNA sequence.
BM480018
BM480018.1 GI:18529060
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1074)
NTN-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12112 row: e column: 13
High quality sequence stop: 652.
Location/Qualifiers
1. 1074

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5491572"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."
BASE COUNT      326 a 256 c 220 g 271 t 1 others
ORIGIN
Query Match      25.4%; Score 877.6; DB 13; Length 1074;
Best Local Similarity 94.4%; Pred. No. 1.4e-182;
Matches 954; Conservative 0; Mismatches 50; Indels 7; Gaps 4;

QY 553 AAAATAAGCGAAAGCTCTTGGCCAGACGATGACATGATTCAGGTAAAGATCTCTA 612
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4 AAAATAGCGAAAGCTCTTGGCCAGACGATGACATGATTCAGGTAAAGATCTCTA 63
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 613 GAAACAGCCAGGGGAAAGAGCTCACCGACCAAGACATGCCAGTGGTACAGGAATGTCA 672
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 GAAACAGCCAGGGGAAAGAGCTCACCGACCAAGACATGCCAGTGGTACAGGAATGTCA 123
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 673 AGGACAGTGTAGTCTAGTGAAGTGAACCTCAACTGTGTACAACTGTCTCTGAAGAA 732
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 AGGACAGTGTAGTCTAGTGAAGTGAACCTCAACTGTGTACAACTGTCTCTGAAGAA 183
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 733 CTCACCTTCTAGAGACAATAGAGACTCCAAGACCTGGGAAACTCTTCCCCAAGATGTAA 792
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 CTCACCTTCTAGAGACAATAGAGACTCCAAGACCTGGGAAACTCTTCCCCAAGATGTAA 243
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 793 CGAGCTCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCGGTGCTGGTAGGA 852
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 CGAGCTCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCGGTGCTGGTAGGA 303
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 853 AAACAAATGAATCTGTAGTGAGCCGCCGAAAGGCTTTATGTATTCAGAAACAATG 912
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 AAACAAATGAATCTGTAGTGAGCCGCCGAAAGGCTTTATGTATTCAGAAACAATG 363
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 913 AAAATCCCTCAGAGTGTTCATCATCATCAAGCTTACTGACATCTCATGGCATGCC 972
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 AAAATCCCTCAGAGTGTTCATCATCATCAAGCTTACTGACATCTCATGGCATGCC 423
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 973 AGGTTCCGCTGAATGCAACAGAGTTCAACTATCTCTCCAGCATCATCAACCAATG 1032
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 AGGTTCCGCTGAATGCAACAGAGTTCAACTATCTCTCCAGCATCATCAACCAATG 483
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1033 ATGCTAGATCTGTCTGATTCATCAAGTGAAGAGGCTGAATCCCTCCAAAGACCT 1092
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 ATGCTAGATCTGTCTGATTCATCAAGTGAAGAGGCTGAATCCCTCCAAAGACCT 543
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1093 ATTCAATTAAGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1152
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 544 ATTCAATTAAGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 603
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1153 CTCTGCTGGGGTATCTTAGTGCCTCTCATGAATCGGTGTTTTCAAATTCCTCTGA 1212
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 604 CTCTGCTGGGGTATCTTAGTGCCTCTCATGAATCGGTGTTTTCAAATTCCTCTGA 663
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1213 GTTTCCTTGTGGCAGTGGCGCTTGGGACCTTCAGTGGTGGTGGTGGTGGTGGT 1272
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 664 GTTTCCTTGTGGCAGTGGCGCTTGGGACCTTCAGTGGTGGTGGTGGTGGTGGT 723
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1273 CACATTCATGCAAGTCAACCAATGATGATGATGATGATGATGATGATGATGATGAT 1332
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 724 CACATTCATGCAAGTCAACCAATGATGATGATGATGATGATGATGATGATGATGAT 783
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1333 AAAGAGGACCACTTTTCACTGATCTCTCTTCAAAACATAGAGAAAGTGCCTATTG 1392
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 784 AAAGAGGACCACTTTTCACTGATCTCTCTTCAAAA-ATTAAAGAAAGTGCCTATTG 842
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1393 ATTCCAGCTGGAGGCTCTAACAGCTCTAGGAGCTCTATTCTTCTTCTTGTGAAC 1452
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 843 ATTCCCGTGGAA-GGTCTAACAGCTCTAGAAGCCGCGCATCTTCTTCTTGTGAAC 901
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1453 ATGCTCTCATCATGTATCAACAATTTAAGATTAAGAAAGAAAGAAAGTCAAGAAACCTG 1512
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 902 ATGCCGCCCATTTGGTC-AACCTTTTGGGTAGGAAGAAAGAAAGATCTCTCAAAAACCTG 960
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1513 AAA----ATCATGATGATGTGGAGATTAAAGAGCAGTGTGTCCCAAGTATCAA 1559
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 AAATGTCATGATGATGTTGGAATTTAAAAAACCGTTTGGCCCCCAATGAA 1011
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
BO716456 915 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8291338 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6194809 5', mRNA sequence.
ACCESSION BO716456
VERSION BO716456.1 GI:21855353
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-re@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Cloned by: Agencourt Bioscience Corporation
Cloned distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13600 row: k column: 02
High quality sequence stop: 590.
Location/Qualifiers
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6194809"
/clone_lib="Lupski_sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACCGCTCCG-3' and
5'-GACTAGTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
BASE COUNT      274 a 118 c 204 g 317 t 2 others
ORIGIN
Query Match      23.1%; Score 800.6; DB 14; Length 915;
Best Local Similarity 96.8%; Pred. No. 1.3e-165;
Matches 880; Conservative 0; Mismatches 21; Indels 8; Gaps 6;

QY 2040 CATGAGTGGCTCATGAATTAGGTGACTTTGCTTCTTACTAAAGCTGGCATGACCGTT 2099
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1 CATGAGTGGCTCATGAATTAGGTGACTTTGCTTCTTACTAAAGCTGGCATGACCGTT 60
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2100 AAGCAGCTCTCTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 2159
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute

TITLE
JOURNAL
COMMENT

FEATURES
source

1. .872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1007185"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18FL3"

BASE COUNT 245 a 187 c 199 g 236 t 5 others
ORIGIN

Query Match 23.1%; Score 799.6; DB 9; Length 872;
Best Local Similarity 96.9%; Pred. No. 2.1e-165;
Matches 834; Conservative 0; Mismatches 24; Indels 3; Gaps 2;

QY 1433 TTTCATGTTTCTGTTGAACATGCTCCTCACATTTGATGATCAACAACTTTAAAGATAAGAGAA 1492
DB 1 TTTCATGTTTCTGTTGAACATGCTCCTCACATTTGATGATCAACAACTTTAAAGATAAGAGAA 60
QY 1493 AAAGATCAGAGAAACCTGAAATGATGATGTTGGAGATTAAAGAGACAGTTGTCCAA 1552
DB 61 AAGAATCAGAGAAACCTGAAATGATGATGTTGGAGATTAAAGAGACAGTTGTCCAA 120
QY 1553 GTATGAATCTCAACTTTCAACAAAGAGAGAAAGTAGATACAGATGATCGAATGAAGG 1612
DB 121 GTATGATCTCAACTTTCAACAAAGAGAGAAAGTAGATACAGATGATCGAATGAAGG 180
QY 1613 CTATTTACGAGCAGACTCAACAGAGCCCTCCACACTTTTGTATCTTCAGACGCCGTGAGCTTT 1672
DB 181 CTATTTACGAGCAGACTCAACAGAGCCCTCCACACTTTTGTATCTTCAGACGCCGTGAGCTTT 240
QY 1673 GGAAGAAGAGAGGTCTATGATGCTCATGCTCATCCACAGGAAGTCTACAAATGAATATGT 1732
DB 241 GGAAGAAGAGAGGTCTATGATGCTCATGCTCATCCACAGGAAGTCTACAAATGAATATGT 300
QY 1733 ACCCAGAGGTGCAAGAAATAAATGCCATTTCACATTTCCACGATACACTCGGCCAGTCAGA 1792
DB 301 ACCCAGAGGTGCAAGAAATAAATGCCATTTCACATTTCCACGATACACTCGGCCAGTCAGA 360
QY 1793 CGATCTCATTACACACCATCATGACTACCATCATATTTCTCCATCATCACCACCACCAAA 1852
DB 361 CGATCTCATTACACACCATCATGACTACCATCATATTTCTCCATCATCACCACCACCAAA 420
QY 1853 CCACCATCTCAGAGTCAACAGCCAGCGCTACTCTCGGAGGAGCTGAAAGATGCCGCGT 1912
DB 421 CCACCATCTCAGAGTCAACAGCCAGCGCTACTCTCGGAGGAGCTGAAAGATGCCGCGT 480
QY 1913 CCCCATTCTGGCCTGGATGGTGAATAATGGGTGATGATGATGATGATGATGATGATGATGATGAT 1972
DB 481 CCCCATTCTGGCCTGGATGGTGAATAATGGGTGATGATGATGATGATGATGATGATGATGATGAT 540
QY 1973 AGCAATTTGGTCTGCTTTTACTGAGGCTTATCAAGTGGTTTAACTACTTCTGTTGCTGT 2032
DB 541 AGCAATTTGGTCTGCTTTTACTGAGGCTTATCAAGTGGTTTAACTACTTCTGTTGCTGT 600
QY 2033 GTTCTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2092

DB 61 AAGCAGGCTGCTCTTTAATAATGCAATGTCACCCATGCTGGGTATCTTGAATGGCAACA 120
QY 2160 GGAATTTTCATTTGTCATGCTCAAAATGTTCTATGTTGGATATTTGACATTTACTGCT 2219
DB 121 GGAATTTTCATTTGTCATGCTCAAAATGTTCTATGTTGGATATTTGACATTTACTGCT 180
QY 2220 GGTCTATTATGATGTTGCTCTGTTGATATGTTGATGTTGATGTTGATGTTGATGTTGAT 2279
DB 181 GGTCTATTATGATGTTGCTCTGTTGATATGTTGATGTTGATGTTGATGTTGATGTTGAT 240
QY 2280 AGTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2339
DB 241 AGTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 2340 GGTGTTGGAATATGTTACTTATTTTCCATATTTTCCATATTTTCCATATTTTCCATATTA 2399
DB 301 GGTGTTGGAATATGTTACTTATTTTCCATATTTTCCATATTTTCCATATTTTCCATATTA 360
QY 2400 TTCTAGTTAAGGTTTAAATGCTAGAGTACGTTTAAAGATGTTGATGTTGATGTTGATGTTGAT 2459
DB 361 TTCTAGTTAAGGTTTAAATGCTAGAGTACGTTTAAAGATGTTGATGTTGATGTTGATGTTGAT 420
QY 2460 TAGGAGATGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2519
DB 421 TAGGAGATGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 2520 TTCTTATTGAATATTTGCTGCTCTTACAAAGTCAGTTAAAGGTCAGTTAAAGGTCAGTTAAAG 2579
DB 481 TTCTTATTGAATATTTGCTGCTCTTACAAAGTCAGTTAAAGGTCAGTTAAAGGTCAGTTAAAG 540
QY 2580 TTATCTATCTTGGAGATAAATCTGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2639
DB 541 TTATCTATCTTGGAGATAAATCTGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 600
QY 2640 GTAACAGAGATTTGGCATGACATGTTCTGATGTTTTCAGGAAATATGTTTAAATGTC 2699
DB 601 GTAACAGAGATTTGGCATGACATGTTCTGATGTTTTCAGGAAATATGTTTAAATGTC 660
QY 2700 TTTTCAAGAACTAACACAGTTATTTCTATGTTTCTGATGTTTCTGATGTTTCTGATGTTTCT 2758
DB 661 TTTTCAAGAACTAACACAGTTATTTCTATGTTTCTGATGTTTCTGATGTTTCTGATGTTTCT 720
QY 2759 GGTGTTTAGGAATAGAAATGTCGCA-TGAAGCCTAAATATACCAAGAAAGCTTATACT-GAA 2816
DB 721 GGTGTTTAGGAATAGAAATGTCGCA-TGAAGCCTAAATATACCAAGAAAGCTTATACT-GAA 780
QY 2817 TTTAAGCAAGAAATAAA-GGAGAAACAGAGAAATCT-GAGAAATTTGGGAGGAGCATAGAT 2874
DB 781 TTTAAGCAAGAAATAAA-GGAGAAACAGAGAAATCT-GAGAAATTTGGGAGGAGCATAGAT 840
QY 2875 TCTTATAAAATCACAAATTTTGTGTAATTT---AGAGGGGAGAAATTTAGAAATTAAGT 2931
DB 841 TCTTATAAAATCCCAAATTTTGTGTAATTTTGTGTAATTTAAGAGGGAAGAAATTTTAAATTAAGT 900

RESULT 3
LOCUS AU120027 HDMBA1 Homo sapiens cDNA clone HEMBA1007185 5', mRNA
DEFINITION AU120027 linear EST 01-AUG-2002

ACCESSION AU120027
VERSION AU120027
KEYWORDS EST

SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 872)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

BASE COUNT

ORIGIN					
	Query Match	20.5%	Score 711:	DB 14:	Length 712;
	Best Local Similarity	99.9%;	Pred. No.	6.6e-146;	
	Matches 711;	Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
QY	2099	TAAGCAGGCTGCCTTTATAAATGCAATTGTACGCCATGTCGGCGTATCTTGGAAATGGCAAC	2158		
Db	1	TAAGCAGGCTGCCTTTATAAATGCAATTGTACGCCATGTCGGCGTATCTTGGAAATGGCAAC	60		
QY	2159	AGGAATTTTCATTTGGTCATTTATGCTCAAATATGTTCTATGTGGATATTTGCACCTACTGTC	2218		
Db	61	AGGAATTTTCATTTGGTCATTTATGCTCAAATATGTTCTATGTGGATATTTGCACCTACTGTC	120		
QY	2219	TGGCTATTTCATGATATGTTGCTCTGTTGATATGATGGTACCTGAAAATGCTCCACAATGATGC	2278		
Db	121	TGGCTATTTCATGATATGTTGCTCTGTTGATATGATGGTACCTGAAAATGCTCCACAATGATGC	180		
QY	2279	TAGTGACCATGGATGTAGCCGCTGGGGGTATTTCTTTTTACAAATGCTGGGATGCTTTT	2338		
Db	181	TAGTGACCATGGATGTAGCCGCTGGGGGTATTTCTTTTTACAAATGCTGGGATGCTTTT	240		
QY	2339	GGGTTTTGGAAATATGTTACTTATTTCCATPATTTGAACATAAAATCGTGTTCGTATATAA	2398		
Db	241	GGGTTTTGGAAATATGTTACTTATTTCCATPATTTGAACATAAAATCGTGTTCGTATATAA	300		
QY	2399	TTTCTAGTTAAGGTTTAAATGCTAGAGTAGCTTAAAAAGTTGTCATAGTTTTCAGTAGGTC	2458		
Db	301	TTTCTAGTTAAGGTTTAAATGCTAGAGTAGCTTAAAAAGTTGTCATAGTTTTCAGTAGGTC	360		
QY	2459	ATAGGGAGATGAGTTGATATGCTGTACTATCGACGCTTTAAAGTTAGTGGGTTTTCTGTAT	2518		
Db	361	ATAGGGAGATGAGTTGATATGCTGTACTATCGACGCTTTAAAGTTAGTGGGTTTTCTGTAT	420		
QY	2519	TTTTGTATTGAATTTGCTGTCTTTACAAAGTCAGTTAAAGTCAGCTTTTAATATTAA	2578		
Db	421	TTTTGTATTGAATTTGCTGTCTTTACAAAGTCAGTTAAAGTCAGCTTTTAATATTAA	480		
QY	2579	GTTATTTCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACAGCTTTATTA	2638		
Db	481	GTTATTTCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACAGCTTTATTA	540		
QY	2639	TGTAACAAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGGAAAAATGCTTTAATG	2698		
Db	541	TGTAACAAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGGAAAAATGCTTTAATG	600		
QY	2699	CTTTTTCGAAGAATTAACACAGTTATTCCTATCTGGAATTTTAGTCTCTGGAAGAACCTGCT	2758		
Db	601	CTTTTTCGAAGAATTAACACAGTTATTCCTATCTGGAATTTTAGTCTCTGGAAGAACCTGCT	660		
QY	2759	GGTGTTTAGGAATTAAGATGTGCATGAAGCTCAAAATACCAAGAAAGCTTAT	2810		
Db	661	GGTGTTTAGGAATTAAGATGTGCATGAAGCTCAAAATACCAAGAAAGCTTAT	712		
RESULT#	7				
LOCUS	BG571701	782 bp	mRNA	linear	EST 10-APR-2001
DEFINITION	602593022P1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4720636 5', mRNA sequence.				
ACCESSION	BG571701				
VERSION	BG571701.1	GI:13579354			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 782)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaabs-r@mail.nih.gov				

```

QY 2654 TGGCATGACATGCTCTGATGTTTCAGGAAAAATGCTTTAATGC-TTTTTCAGAACT 2712
Db 597 TGGCATGACATGCTCTGATGTTTCAGGAAAAATGCTTTAATGC-TTTTTCAGAACT 656

QY 2713 AACACAGTTATTCCTATCTACTGATTTTAGTCTCTGAGAACTGCTGCTGTTTAGGAATA 2772
Db 657 AACACAGTTATTCCTATCTACTGATTTTAGTCTCTGAGAACTGCTGCTGTTTAGGAATA 716

QY 2773 AGAATGTCATGAAGCCCTAAAAA---TACCAGAAAGCTTATCTACTGAATTTTAAGCAAGAA 2829
Db 717 AGAATGTCGTTGAGGCTTAAATTTACCAGAAAGGCTTATCTACTGAATTTTAAGCAAGAA 776

QY 2830 ATAAAG 2835
Db 777 ATAAAG 782

RESULT 8
LOCUS BQ027619/c
DEFINITION UI-H-CO0-aqz-a-11-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone
IMAGE:3105668 3', mRNA sequence.
ACCESSION BQ027619
VERSION BQ027619.1 GI:19762898
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 701)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TUMOR Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-re@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1. 701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3105668"
/clone_lib="NCI_CGAP_Sub9"
/tissue_type="mixed"
/dev_stage="mixed"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p7T3-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoR I; Site_2: Not I; tissues:
Cholonic mucosa with Crohns disease, Cholonic mucosa with
ulcerative colitis, Fetal thymus, Cervix, Cervical
adenosquamous carcinoma, Ligament cells, Prostate
carcinoma, Bladder carcinoma, Brain oligodendrocyte;
NCI_CGAP_Sub9 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are CGTC, AACG, GGCC,
GGAAG, TAGC, TAAG, ATGG, AGACA, ATCAC. For additional
information, contact: Bento Soares, Bento-soares@uiowa.edu
TAG_LIB=UI-H-CO0

```

```

TAG_TISSUE=Prostate Carcinoma
TAG_SEQ=ATGG*
BASE COUNT 198 a 143 c 152 g 208 t
ORIGIN
Query Match 20.0%; Score 692; DB 14; Length 701;
Best Local Similarity 99.3%; Pred. No. 1e-141;
Matches 695; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 805 CACCAGTGTACATCAAAAGAGCGGGTGAGCGCGGTGCTGTAGTAAACAAATGAAT 864
Db 701 CACCAGTGTACATCAAAAGAGCGGGTGAGCGCGGTGCTGTAGTAAACAAATGAAT 642

QY 865 CTGTGAGTGAGCCCGGAAAGGCTTTATGTTATTCAGAAACACAAATGAAATCCCTCAGG 924
Db 641 CTGTGAGTGAGCCCGGAAAGGCTTTATGTTATTCAGAAACACAAATGAAATCCCTCAGG 582

QY 925 AGTGTTCATATGATCAAAAGCTACTGACATCTCATGGCATGGGCATCCAGGTTCCGCTGA 984
Db 581 AGTGTTCATATGATCAAAAGCTACTGACATCTCATGGCATGGGCATCCAGGTTCCGCTGA 522

QY 985 ATGCAACAGAGTTCAACTATCTCTGCCAGCCATCATCAACCAAAATGTAGCTAGATCTT 1044
Db 521 ATGCAACAGAGTTCAACTATCTCTGCCAGCCATCATCAACCAAAATGTAGCTAGATCTT 462

QY 1045 GTCTGATTTCATACAAAGTGAAGAGGCTGAAATCCCTCCAAAGACCTATTCTATTACAAA 1104
Db 461 GTCTGATTTCATACAAAGTGAAGAGGCTGAAATCCCTCCAAAGACCTATTCTATTACAAA 402

QY 1105 TAGCCTGGGTTGGTGGTTTATAGCCATTTCCATCATCAGTTTCTCTCTCTCTCTCTGGGG 1164
Db 401 TAGCCTGGGTTGGTGGTTTATAGCCATTTCCATCATCAGTTTCTCTCTCTCTCTGGGG 342

QY 1165 TTATCTAGTGCCTCTCATCAATCGGGTGTTTTCAAATTTCTCCTGAGTTTCTCTGTGG 1224
Db 341 TTATCTAGTGCCTCTCATCAATCGGGTGTTTTCAAATTTCTCCTGAGTTTCTCTGTGG 282

QY 1225 CACTGGCCGTTGGGACTTTGAGTGGTGTATGCTTTTACACCTTCTTCCACATTTCTCATG 1284
Db 281 CACTGGCCGTTGGGACTTTGAGTGGTGTATGCTTTTACACCTTCTTCCACATTTCTCATG 222

QY 1285 CAAGTCACCATATGATGTCATAGCCATGAAGACAGCAATGGAATGAAAGAGGACCAC 1344
Db 221 CAAGTCACCATATGATGTCATAGCCATGAAGACAGCAATGGAATGAAAGAGGACCAC 162

QY 1345 TTTTCAGTCTATCTGCTTCTCAAAACATAGAAGAAAGTGCCTATTTTGTTCACCGTGA 1404
Db 161 TTTTCAGTCTATCTGCTTCTCAAAACATAGAAGAAAGTGCCTATTTTGTTCACCGTGA 102

QY 1405 AGGCTTAACAGCTCTAGGAGGCGCTGTATTTCATGTTTCTTGTGAACATGTCTCCTCAG 1464
Db 101 AGGCTTAACAGCTCTAGGAGGCGCTGTATTTCATGTTTCTTGTGAACATGTCTCCTCAG 42

QY 1465 TGATCAACAAATTTAAGATGAAGAAAGAAATCAGAA 1504
Db 41 TGATCAACAAATTTAAGATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2

RESULT 9
LOCUS BQ03446
DEFINITION 602419134F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4526145 5',
mRNA sequence.
ACCESSION BQ03446
VERSION BQ03446.1 GI:13296894
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 728)
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10433 row: c column: 10
High quality sequence stop: 723.
Location/Qualifiers
1. 728

FEATURES
source

BASE COUNT 189 a 100 c 163 g 276 t

Note: this is a NIH_MGC Library.
full-length clones and constructed by Life Technologies.

Average insert size 1.7 kb. Library enriched for
Site 2: SalI; Cloned unidirectionally: oligo-dT primed.
/note="Organ: bladder; Vector: pCMV-SPORT6; Site 1: NotI;
/lab_host="DH10B (phage-resistant)"
/tissue_type="transitional cell papilloma, cell line"
/clone_lib="NIH_MGC_93"
/db_xref="IMAGE:4526145"
/organism="Homo sapiens"
/db_xref="taxon:9606"

Query Match 19.6%; Score 676.8; DB 12; Length 728;
Best Local Similarity 98.9%; Pred. No. 2.3e-138;
Matches 724; Conservative 0; Mismatches 2; Indels 6; Gaps 4;

QY 2003 ATCAAGTGGTTAAAGTACTTCTGTTGCTGTTCTGTCATGAGTGGCTCATGATTAAGG 2062
DB 1 ATCAAGTGGTTAAAGTACTTCTGTTGCTGTTCTGTCATGAGTGGCTCATGATTAAGG 60
QY 2063 TGACTTTGCTGTTCTACTAAAGCGTGGCATGACCGTTAAGCAGCGTGTCCTTTAATAGC 2122
DB 61 TGACTTTGCTGTTCTACTAAAGCGTGGCATGACCGTTAAGCAGCGTGTCCTTTAATAGC 120
QY 2123 ATTGTGACGCATGCTGGCGTATCTTGGAAATGCAACAGGAATTTTCATTTGTCATTAAGC 2182
DB 121 ATTGTGACGCATGCTGGCGTATCTTGGAAATGCAACAGGAATTTTCATTTGTCATTAAGC 180
QY 2183 TGAATAATGTTCTATGTGGATATTTGCACCTACTGCTGGCTTATTCATGTAATGTGCTCT 2242
DB 181 TGAATAATGTTCTATGTGGATATTTGCACCTACTGCTGGCTTATTCATGTAATGTGCTCT 240
QY 2243 GGTGATATGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGATGTAGCCGCTG 2302
DB 241 GGTGATATGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGATGTAGCCGCTG 300
QY 2303 GGGGTATTTCTTTTACAGAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTAT 2362
DB 301 GGGGTATTTCTTTTACAGAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTAT 358
QY 2363 TTCCATATTTGAACATAAAATCGTGTTCGTATAAAATTCCTAGTTAAGGTTTAAATGCTA 2422
DB 359 TTCCATATTTGAACATAAAATCGTGTTCGTATAAAATTCCTAGTTAAGGTTTAAATGCTA 418
QY 2423 GAGTAGCTTTAAAGCTTGTCTACATTTTCAGTAGGTCATAGGAGATGATTTGTATGCTG 2482
DB 419 GAGTAGCTTTAAAGCTTGTCTACATTTTCAGTAGGTCATAGGAGATGATTTGTATGCTG 477
QY 2483 TACTATGACCGCTTTAAAGCTTGTGATTTTGTGATTTTGTATTTGATATTTGCTGCTG 2542
DB 478 TACTATGACCGCTTTAAAGCTTGTGATTTTGTGATTTTGTATTTGATATTTGCTGCTG 537
QY 2543 TTACAAGTCAGTTAAAGTCAGTTTAAATATTTAAGTATTTCTATCTTGGAGATAAAT 2602
DB 538 TTACAAGTCAGTTAAAGTCAGTTTAAATATTTAAGTATTTCTATCTTGGAGATAAAT 597
QY 2603 CTGTATGTCGAATTCACCGGTATTACCAGTTTATTATGTAACAAGACATTTGGCATCAC 2662

DB 598 CTGTATGTCGAATTCACCGGTATTACCAGTTTATTATGTAACAAGACATTTTGGATGAC 655
QY 2663 ATGTTCTGTATGTTTTCAGGG-AAAAATGCTTTTAATGCTTTTTCAGAGAACTAACACAGTT 2721
DB 656 ATGTTCTGTATGTTTTCAGGGAAAAAATGCTTTTAATGCTTTTTCAGAGAACTAACACAGTT 715
QY 2722 ATTCCTATACCTG 2733
DB 716 ATTCCTATACCTG 727

RESULT 10
BI222452 830 bp mRNA linear EST 11-JUL-2001
LOCUS 602940041F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5102890 5',
DEFINITION mRNA sequence.
ACCESSION BI222452
VERSION BI222452.1 GI:4675896
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 830)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11248 row: b column: 11
High quality sequence stop: 765.
Location/Qualifiers
1. 830
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:5102890"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 256 a 100 c 183 g 291 t

ORIGIN

Query Match 19.1%; Score 661.6; DB 13; Length 830;
Best Local Similarity 93.6%; Pred. No. 5.1e-135;
Matches 769; Conservative 0; Mismatches 39; Indels 14; Gaps 7;

QY 2186 AAATGTTTCTATGTGGATATTTGCACCTACTGCTGGCTTATTCATGTTGCTCTGG- 2244
DB 2 AAATGTTTCTATGTGGATATTTGCACCTACTGCTGGCTTATTCATGTTGCTCTGGC 61
QY 2245 TTGATATGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGTAGCCGCTGG 2304
DB 62 TTGATATGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGTAGCCGCTGG 121
QY 2305 GGTATTTCTTTTACACAATGCTGGGATGCTTTTGGTGGTGGAAATGATGTTACTTATTT 2364
DB 122 GGTATTTCTTTTACACAATGCTGGGATGCTTTTGGTGGTGGAAATGATGTTACTTATTT 179
QY 2365 CCATATTTGAACATAAAATCGTGTTCGTATAAAATTTCTAGTTAAGGTTTAAATGCTAGA 2424
DB 180 CCATATTTGAACATAAAATCGTGTTCGTATAAAATTTCTAGTTAAGGTTTAAATGCTAGA 239

QY	2425	GTAGCTTAAAAAGTTGTTCATAGTTTCAGTAGGTCAATAGGGAGATGAGTTTGTATGCTGTGA	2484
DB	240	GTAGCTTAAAAAGTTGTTCATAGTTTCAGTAGGTCAATAGGGAGATGAGTTTGTATGCTGTGA	299
QY	2485	CTATGACAGCGTTTAAAGTTAGTGGGTTTGTGATTTTTTGTATTGAATATTCGTCCTGTGT	2544
DB	300	CTATGACAGCGTTTAAAGTTAGTGGGTTT - GTGATTTTTTGTATTGAATATTCGTCCTGTGT	357
QY	2545	ACAACTCAGTTTAAAGGTACGTTTTAAATATTAACTTATCTATCTCTGGAGATAAAATCT	2604
DB	358	ACAACTCAGTTTAAAGGTACGTTTTAAATATTAACTTATCTATCTCTGGAGATAAAATCT	417
QY	2605	GTATGTGCAATTCACCGGTATTACCAAGTTTATTATGTAACAAGAGATTTGGCATGACAT	2664
DB	418	GTATGTGCAATTCACCGGTATTACCAAGTTTATTATGTAACAAGAGATTTGGCATGACAT	477
QY	2665	GTTCTCTATGTTTCAGGGAAAAATGCTTTAATGCTTTTTTCAAGAACTTAACACAGTTATT	2724
DB	478	GTTCTCTATGTTTCAGGGAAAAATGCTTTAATGCTTTTTTCAAGAACTTAACACAGTTATT	537
QY	2725	CCATATCTCGATTTTAAAGTCTCTGAAGAACTCTCGTGTTTAGGAATTAAGAACTGTGCATG	2784
DB	538	CCATATCTCGATTTTAAAGTCTCTGAAGAACTCTCGTGTTTAGGAATTAAGAACTGTGCATG	597
QY	2785	AAGCCTTAAATACCAAGAAAGCTTTACTTGAATTTTAAAGCAAGAAATAAAGGAGAAAAAGA	2844
DB	598	AAGCCTTAAATACCAAGAAAGCTTTACTTGAATTTTAAAGCAAGAAATAAAGGAGAAAAAGA	657
QY	2845	GAGAAATCTGAGAAATTTGGGAGGCATAGATTCCTTAT - AAAAATCACAAAATTTGTGTAA	2903
DB	658	GAGAAATCTGAGAA - TGGGAGGCATAGATTCCTTATTAATAAATCACAAAATTTGTGTAC	716
QY	2904	ATTAGAGGGGAGAAATTT - - AGAATTAAGTATATAAAAGGCGAAGTATAGTATAGAGTACA	2966
DB	717	ATTAGAGGGGAGAAATTTACGAATTAAGTTTATTAAAGGCGGAATTAGTATAGAGTTACG	776
QY	2961	TTCAATTAACATTTTGTTCAGGATTTATTCCTCGTAAAAACGT	3002
DB	777	TTCTTAAAA - -TTTGCCAGGATTTATTCCTCGTAAAAAGTAGT	814
RESULT 11	AL039253	697 bp mRNA linear EST 29-FEB-2001	
LOCUS	DKFZp727K071_r1 727 (synonym: hmcfl)	Homo sapiens cDNA clone	
DEFINITION	DKFZp727K071 5', mRNA sequence.		
ACCESSION	AL039253		
VERSION	AL039253.1	GI:5408327	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 697)		
TITLE	Poustka A., Klein M., Mewes H.W., Gassenhuber J. and Wiemann S.		
JOURNAL	EST (Poustka, et al.)		
COMMENT	Unpublished (1999)		
	Contact: Poustka A.J.		
	Department Lehrfach		
	Max-Planck-Institute for Molecular Genetics		
	Innistrasse 73, 14195 Berlin, Germany		
	Tel.: +49-30-84131623		
	Fax: +49-30-84131128		
	Email: poustka@mpg-berlin-dahlem.mpg.de		
	This is the 5' sequence of the clone insert		
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;		
	sequenced by DKFZ (German Cancer Research Center,		
	Heidelberg/Germany) within the cDNA sequencing consortium of the		
	German Genome Project.		
	No sl sequence available.		
	This clone (DKFZp727K071) is available at the RZPD in Berlin.		

FEATURES		source		Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonegrzpd.de.	
Location/Qualifiers		1. .697			
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="DKFZp727K071"					
/clone_lib="727 (synonym: hmcfl)"					
/tissue_type="breast cancer"					
/dev_stage="adult"					
/lab_host="DH10B"					
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"					
BASE COUNT		247 a	76 c	136 g	238 t
ORIGIN					
Query Match		18.9%; Score 652.6; DB 9; Length 697;			
Best Local Similarity		98.9%; Pred. No. 4.8e-133;			
Matches 689; Conservative		0;	Mismatches	4;	Gaps
Qy	2411	GTTTAAATGCTAGAGTAGCTTAAAAAGTTGTCATAGTTTCAGTAGGTGCATAGGAGATGA	2470		
Db	1	GTTTAAATGCTAGAGTAGCTTAAAAAGTTGTCATAGTTTCAGTAGGTGCATAGGAGATGA	60		
Qy	2471	GTTTGTATGCTGTACTATACGCGGTTTAAAGTTAGTGGGTTTTGTGATTTTTTGTATTGAA	2530		
Db	61	GTTTGTATGCTGTACTATACGCGGTTTAAAGTTAGTGGGTTTTGTGATTTTTTGTATTGAA	120		
Qy	2531	TATTGCTGTCGTTCACAAAGTCAGTTTAAAGGTACGTTTTTAAATATTAAAGTTATTCTATCT	2590		
Db	121	TATTGCTGTCGTTCACAAAGTCAGTTTAAAGGTACGTTTTTAAATATTAAAGTTATTCTATCT	180		
Qy	2591	TGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAGTTTATTATGTAAACAAGAG	2650		
Db	181	TGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAGTTTATTATGTAAACAAGAG	240		
Qy	2651	ATTGGCATGACATGTTCTGTATGTTTCAGGGAAAAATGTCCTTTAATGCTTTTTCAGAA	2710		
Db	241	ATTGGCATGACATGTTCTGTATGTTTCAGGGAAAAATGTCCTTTAATGCTTTTTCAGAA	300		
Qy	2711	CTAACACAGTTATTCTCTATACTGAGTTTATAGTCTCTGAAGAACTGCCTGGTGTTAGGAA	2770		
Db	301	CTAACACAGTTATTCTCTATACTGAGTTTATAGTCTCTGAAGAACTGCCTGGTGTTAGGAA	360		
Qy	2771	TAAGAATGTGCATGAAGCCTAAATACCAGAAAGCTTATCTAGAAATTTAAGCAAGAAA	2830		
Db	361	TAAGAATGTGCATGAAGCCTAAATACCAGAAAGCTTATCTAGAAATTTAAGCAAGAAA	420		
Qy	2831	TAAAGGAGAAAAGAGAGAATCTGACAATTGGGAGGCGATAGATTCTTTATAAAAATCACA	2890		
Db	421	TAAAGGAGAAAAGAGAGAATCTGAGAATTTGGGAGGCGATAGATTCTTTATAAAAATCACA	480		
Qy	2891	AAATTTCGTTCTAAATTAGAGGGGAGAAATTTAGAAATTAAGTATAAAAAGGCGAGAATTAGT	2950		
Db	481	AAATTTCGTTCTAAATTAGAGGGGAGAAATTTAGAAATTAAGTATAAAAAGGCGAGAATTAGT	540		
Qy	2951	ATAGAGTACATTCATTAAACATTTTTTGTCCAGGATTTATTTCCCGTAAAAACGCTAGTGAGCA	3010		
Db	541	ATAGAGTACATTCATTAAACATTTTTTGTCCAGGATTTATTTCCCGTAAAAACGCTAGTGAGCA	600		
Qy	3011	C-TCCTATATCTAATT--AGTGTACATTTAACTTTGTATATATACAGAATCTTAATATA	3061		
Db	601	CTTTTCATATCTAATTTAGTTTGTACATTTAACTTTGTATATATACAGAATCTTAATATA	660		
Qy	3068	TTTAAATGAATTCACAGCAATATA-CACTTTGACCAAGAA	3103		
Db	661	TTTAAATGAATTCACAGCAATATATACCTTGACCAAGAA	697		
RESULT 12					
BM560789		1135 bp		mrna	linear
LOCUS		AGENCOURT_65660.13		NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5550	
DEFINITION		5', mRNA sequence.			

ACCESSION BM560789
VERSION BM560789.1 GI:18805500
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1135)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M12263 row: e column: 16
High quality sequence stop: 423.
Location/Qualifiers
1..1135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5550711"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 Kb. Library constructed by Life
Technologies."
BASE COUNT 292 a 255 c 266 g 322 t
ORIGIN
Query Match 18.8%; Score 652; DB 13; Length 1135;
Best Local Similarity 98.3%; Pred. No. 6.7e-133;
Matches 680; Conservative 0; Mismatches 10; Indels 2; Gaps 2;
QY 1601 TCGAAGCTGAAGGCTATTTACGAGCAGACTCACAGAGCCCTCCACCTTGATTTCTCAGCA 1660
Db 2 TCTCAATGAAGGCTATTTACGAGCAGACTCACAGAGCCCTCCACCTTGATTTCTCAGCA 61
QY 1661 GCCTGCAGCTTTCGAGAGAGAGGCTCATGATAGCTCATGCTCATCCACAGGAAGTCTA 1720
Db 62 GCCTGCAGCTTTCGAGAGAGAGAGGCTCATGATAGCTCATGCTCATCCACAGGAAGTCTA 121
QY 1721 CAATGAATATGTACCCAGAGGGTGCAAGAAATAATGCCATTACACATTTCCAGGATACACT 1780
Db 122 CAATGAATATGTACCCAGAGGGTGCAAGAAATAATGCCATTACACATTTCCAGGATACACT 181
QY 1781 CGGCCAGTCAGAGATCTCATTCACCCACCATCATGACTACCATCATATATTCCTCCATCATCA 1840
Db 182 CGGCCAGTCAGAGATCTCATTCACCCACCATCATGACTACCATCATATATTCCTCCATCATCA 241
QY 1841 CCACCAACCAACCAACCATCTCCACAGTCACAGCCAGCCGCTACTCTCGGGAGGAGCTGAA 1900
Db 242 CCACCAACCAACCAACCATCTCCACAGTCACAGCCAGCCGCTACTCTCGGGAGGAGCTGAA 301
QY 1901 AGATGCGGGGTCGCCACTTTGGCCCTGGATGGTAAATGGGTGATGGCTGCACAAATTT 1960
Db 302 AGATGCGGGGTCGCCACTTTGGCCCTGGATGGTAAATGGGTGATGGCTGCACAAATTT 361
QY 1961 CAGCGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAGTAC 2020
Db 362 CAGCGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAGTAC 421
QY 2021 TTCTGTGCTGTGTTCTCATGATGTTGCCTCATGATAGTGGTACTTTGCTTCTACT 2080
Db 422 TTCTGTGCTGTGTTCTCATGATGTTGCCTCATGATAGTGGTACTTTGCTTCTACT 481

QY 2081 AAAGCTGCATGACCGTTAAGCAGGCTGCTTTATATCATTTGTGACGCAATGCTGGC 2140
Db 482 AAAGCTGCATGACCGTTAAGCAGGCTGCTTTATATCATTTGTGACGCAATGCTGGC 541
QY 2141 GTATCTTGAATGGCAACAGGAATTTTCATTTGGTCAATTCATGCTGAAATGTTTCTATGTG 2200
Db 542 GTATCTTGAATGGCAACAGGAATTTTCATTTGGTCAATTCATGCTGAAATGTTTCTATGTG 601
QY 2201 GATATTTGCACTTACCTGCTGGCTTATTCATGCTATGCTCTGCTGTTGATATGGTA-CCTG 2259
Db 602 GATATTTGCACTTACCTGCTGGCTTATTCATGCTATGCTCTGCTGTTGATATGGGACCTG 661
QY 2260 AAATGCTGCACAA-TGATGCTAGTGACCATGG 2290
Db 562 AAATGCTGCACAAATTCATGCTAATGAACATGG 693
RESULT 13
BM978669/c
LOCUS
DEFINITION
BM978669 674 bp mRNA linear EST 21-MAR-2002
UI-CF-DUI-ads-1-15-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
UI-CF-DUI-ads-1-15-0-UI 3', mRNA sequence.
ACCESSION BM978669.1 GI:19598338
VERSION BM978669
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 674)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-23, >AT-rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-ads-1-15-0-UI"
/clone_lib="UI-CF-DUI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into p773-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2002, 17:38:12 ; Search time 96.4981 Seconds
(without alignments)
10999.262 Million cell updates/sec

Title: US-09-642-034-1
Perfect score: 3461
Sequence: 1 CTCGTGCCGAATCCGCCG.....AAATTCCTATATCAGCTTG 3461

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

searched: 441362 seqs, 153338381 residues

total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

```
Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	2249	65.0	2404	1	US-08-311-023-3		Sequence 3, Appl
2	1298	37.2	1310	1	US-08-311-023-1		Sequence 1, Appl
3	360	10.4	385	4	US-09-712-016-78		Sequence 78, Appl
c 4	165	4.8	177	4	US-09-020-956-162		Sequence 162, App
c 5	165	4.8	177	4	US-09-020-607-162		Sequence 162, App
c 6	165	4.8	177	4	US-09-603-785-162		Sequence 162, App
c 7	165	4.8	177	4	US-09-439-313-162		Sequence 162, App
c 8	165	4.8	177	4	US-09-352-616A-162		Sequence 162, App
c 9	165	4.8	177	4	US-09-232-149A-162		Sequence 64, Appl
c 10	59.8	1.7	285	4	US-08-990-571-64		Sequence 64, Appl
c 11	59.8	1.7	285	4	US-09-528-784A-64		Sequence 65, Appl
c 12	59.8	1.7	342	4	US-08-990-571-65		Sequence 65, Appl
c 13	59.8	1.7	342	4	US-09-528-784A-65		Sequence 60, Appl
c 14	59.8	1.7	351	4	US-08-990-571-60		Sequence 60, Appl
c 15	59.8	1.7	351	4	US-09-528-784A-60		Sequence 63, Appl
c 16	59.8	1.7	356	4	US-08-990-571-63		Sequence 63, Appl
c 17	59.8	1.7	356	4	US-09-528-784A-63		Sequence 58, Appl
c 18	59.8	1.7	358	4	US-08-990-571-58		Sequence 58, Appl
c 19	59.8	1.7	358	4	US-09-528-784A-58		Sequence 66, Appl
c 20	59.8	1.7	363	4	US-08-990-571-66		Sequence 67, Appl
c 21	59.8	1.7	363	4	US-08-990-571-67		Sequence 66, Appl
c 22	59.8	1.7	363	4	US-09-528-784A-66		Sequence 67, Appl
c 23	59.8	1.7	363	4	US-09-528-784A-67		Sequence 59, Appl
c 24	59.8	1.7	409	4	US-08-990-571-59		Sequence 59, Appl
c 25	59.8	1.7	409	4	US-09-528-784A-59		Sequence 61, Appl
c 26	59.8	1.7	410	4	US-08-990-571-61		Sequence 61, Appl
c 27	59.8	1.7	410	4	US-09-528-784A-61		Sequence 61, Appl

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match      65.0%; Score 2249; DB 1; Length 2404;
Best Local Similarity 97.4%; pred. No. 0;
Matches 2341; Conservative 0; Mismatches 0; Indels 62; Gaps 3;
QY 1119 GCTTTTATAGCCATTCCATCATCAGTTTCTCTCTGCTGGGGGTATCTATTAGTGCCT 1178

```

Db 1 GGTATTATAGCCATTTCCATCATCAGTTTCCCTGTCTCTGCTGGGGTTATCTAGTGCCT 60
Qy 1179 CTCATGAATCGGGTGTATTTTCAAAATTTCTCCTGAGTTTCTCTGGCAGCTGGCCGTGG 1238
Db 61 CTCATGAATCGGGTGTATTTTCAAAATTTCTCCTGAGTTTCTCTGGCAGCTGGCCGTGG 120
Qy 1239 ACTTTGAGTGGTGATGCTTTTTCACCTTCTTCCACATTTCTCATGCAAGTCAACCAAT 1298
Db 121 ACTTTGAGTGGTGATGCTTTTTCACCTTCTTCCACATTTCTCATGCAAGTCAACCAAT 180
Qy 1299 AGTCATAGCCATGAAGAACCCAGCAATGGAAATGAAGAGAGACACATTTTCAGTCATCTG 1358
Db 181 AGTCATAGCCATGAAGAACCCAGCAATGGAAATGAAGAGAGACACATTTTCAGTCATCTG 240
Qy 1359 TCTTCTCAAAACATAGAAGAAAGTGCCTATTTTGATTTCCAGCTGGAAGGGTCTAACAGCT 1418
Db 241 TCTTCTCAAAACATAGAAGAAAGTGCCTATTTTGATTTCCAGCTGGAAGGGTCTAACAGCT 300
Qy 1419 CTAGGAGGCTGTATTTTCATGTTTCTTGTGTGAACATGCTCTCACATGTGATCAACAATTT 1478
Db 301 CTAGGAGGCTGTATTTTCATGTTTCTTGTGTGAACATGCTCTCACATGTGATCAACAATTT 360
Qy 1479 AAAGATAAGAAAGAAAGAAATCAGAGAAACCTGAAATGATGATGATGTGGAGATTAAAG 1538
Db 361 AAAGATAAGAAAGAAAGAAATCAGAGAAACCTGAAATGATGATGATGTGGAGATTAAAG 420
Qy 1539 AAGCAGTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATGAGAGAAAGTAGATACAGAT 1598
Db 421 AAGCAGTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATGAGAGAAAGTAGATACAGAT 480
Qy 1599 GATCGAACTGAAGCTATTTACGAGCAGACTCACAGAGCCCTGCCACATTTTGATTTCTCAG 1658
Db 481 GATCGAACTGAAGCTATTTACGAGCAGACTCACAGAGCCCTGCCACATTTTGATTTCTCAG 540
Qy 1659 CAGCCTGAGTCTTGGGAAGAGAGAGTTCATGATAGTCTATGCTCATCCACAGGAAGTC 1718
Db 541 CAGCCTGAGTCTTGGGAAGAGAGAGTTCATGATAGTCTATGCTCATCCACAGGAAGTC 600
Qy 1719 TACAATGAATATGTACCCAGAGGGTGCAGAAATATAATGCCATTCACATTTCCACGATACA 1778
Db 601 TACAATGAATATGTACCCAGAGGGTGCAGAAATATAATGCCATTCACATTTCCACGATACA 660
Qy 1779 CTCGCCAGTCAAGCATCTCATTCACACCATCATGACTACCATCATATTTCCATCAT 1838
Db 661 CTCGCCAGTCAAGCATCTCATTCACACCATCATGACTACCATCATATTTCCATCAT 720
Qy 1839 CACCACCCAAACACCATCTCTCACAGTCAAGCCAGCCCTACTCTCGGGAGGAGCTG 1898
Db 721 CACCACCCAAACACCATCTCTCACAGTCAAGCCAGCCCTACTCTCGGGAGGAGCTG 780
Qy 1899 AAGATGCCGGCTGCCACTTTGGCCTGGATGGTGAATAGGTGATGGCTGCACAT 1958
Db 781 AAGATGCCGGCTGCCACTTTGGCCTGGATGGTGAATAGGTGATGGCTGCACAT 840
Qy 1959 TTCAGCGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGCTTATCAAGTGGTTTAAGT 2018
Db 841 TTCAGCGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGCTTATCAAGTGGTTTAAGT 900
Qy 2019 ACTTCTGCTGCTGTCTGTCTATGATGATGGCTTCATGAATAGGTGACTTTGCTGTTCTA 2078
Db 901 ACTTCTGCTGCTGTCTGTCTATGATGATGGCTTCATGAATAGGTGACTTTGCTGTTCTA 960
Qy 2079 CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTTATTAATGCAATGTGCAGCCATGCTG 2138
Db 961 CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTTATTAATGCAATGTGCAGCCATGCTG 1020
Qy 2139 CGGTATCTTGAATGGCAACAGGAATTTTTCATTTGGTTCATATGCTGAAAAATGTTCTATG 2198
Db 1021 CGGTATCTTGAATGGCAACAGGAATTTTTCATTTGGTTCATATGCTGAAAAATGTTCTATG 1080
Qy 2199 TGGATATTTGCACCTTACTGCTGGCTATTCATGATGATGCTGCTGCTGATATGGTACCT 2258
Db 1081 TGGATATTTGCACCTTACTGCTGGCTATTCATGATGATGCTGCTGCTGATATGGTACCT 1140

Qy 2259 GAAATGCTGCACAAATGATGCTAGTGACCATGGATGAGCGCTGGGGTATTTCTTTT 2318
Db 1141 GAAATGCTGCACAAATGATGCTAGTGACCATGGATGAGCGCTGGGGTATTTCTTTT 1200
Qy 2319 CAGAAATGCTGGGATGCTTTTGGGTTTGGAAATATGATTTACTTATTTCCATATTTGAACAT 2378
Db 1201 CAGAAATGCTGGGATGCTTTTGGGTTTGGAAATATGATTTACTTA--TTCCATATTTGAACAT 1259
Qy 2379 AAATFCGTGTTTCTGATATAAATTTCTAGTTAAGGTTTAAATGCTAGAGTAGCTTTAAAAAGT 2438
Db 1260 AAATFCGTG--TTCGTATAAATTTCTAGTTAAGGTTTAAATGCTAGAGTAGCTTTAAAAAGT 1318
Qy 2439 TCTCATAGTTTTCAGTAGCTCATAGGAGATGAGTTTCTATGCTGCTACTATGCGAGCGTTTA 2498
Db 1319 TCTCATAGTTTTCAGTAGCTCATAGGAGATGAGTTTCTATGCTGCTACTATGCGAGCGTTTA 1378
Qy 2499 AAGTATAGTGGGTTTGTGATTTTGTATTTGATTTTGTATTTGCTGCTGTTACAAAGTCAGTTAA 2558
Db 1379 AAGTATAGTGGGTTTGTGATTTTGTATTTGATTTTGTATTTGCTGCTGTTACAAAGTCAGTTAA 1438
Qy 2559 AGTACGTTTAAATATTTAAAGTTATTTCTATCTTGGAGATAAAATCTGATGTCGAATTCA 2618
Db 1439 AGTACGTTTAAATATTTAAAGTTATTTCTATCTTGGAGATAAAATCTGATGTCGAATTCA 1498
Qy 2619 CCGGTATACCCAGTTTATTTATGTAACACAGATTTGGCATGCACATGTTCTGATGTTTC 2678
Db 1499 CCGGTATACCCAGTTTATTTATGTAACACAGATTTGGCATGCACATGTTCTGATGTTTC 1558
Qy 2679 AGGAAAAATGCTTAAATGCTTTTTCAGAACTAACACAGATTTATTCCTATACCTGATTT 2738
Db 1559 AGGAAAAATGCTTAAATGCTTTTTCAGAACTAACACAGATTTATTCCTATACCTGATTT 1618
Qy 2739 TAGTCTCTGAAGAACTGCTGTTTGTAGGAATAAGATGTCATGAAGCCTTAAATATACC 2798
Db 1619 TAGTCTCTGAAGAACTGCTGTTTGTAGGAATAAGATGTCATGAAGCCTTAAATATACC 1678
Qy 2799 AAGAAAGCTTATCTGAAATTTAAGCAAGAAATTAAGGAGAAAAAGAGAAATCTGAGAA 2858
Db 1679 AAGAAAGCTTATCTGAAATTTAAGCAAGAAATTAAGGAGAAAAAGAGAAATCTGAGAA 1738
Qy 2859 TTGGGAGGCAATGATTTTATATAAAATCACAATTTTGTGTTAAATAGAGGGGAGAAA 2918
Db 1739 TTGGGAGGCAATGATTTTATATAAAATCACAATTTTGTGTTAAATAGAGGGGAGAAA 1798
Qy 2919 TTTAGAAATTAAGTATAAAGGCGAGAAATAGTATAGAGTACATTCATTAACATTTTGT 2978
Db 1799 TTTAGAAATTAAGTATAAAGGCGAGAAATAGTATAGAGTACATTCATTAACATTTTGT 1858
Qy 2979 CAGGATTTATTTCCCTGTAAGGCTAGTGCAGCCTCTCATATACTAATTAGTGTACATTTA 3038
Db 1859 CAGGATTTATTTCCCTGTAAGGCTAGTGCAGCCTCTCATATACTAATTAGTGTACATTTA 1918
Qy 3039 ACTTTGTAATAACAGAAATCTAAATATATTTAATGAATTCAGCAATATACACTTGACC 3098
Db 1919 ACTTTGTAATAACAGAAATCTAAATATATTTAATGAATTCAGCAATATACACTTGACC 1978
Qy 3099 AAGAAATTTGGAATTTTCAAAATGTTGCGGGGTTATATACCAGATGAGTACAGTGAAGTAG 3158
Db 1979 AAGAAATTTGGAATTTTCAAAATGTTGCGGGGTTATATACCAGATGAGTACAGTGAAGTAG 2038
Qy 3159 TTTATGTAATCACCAGACTGGGTTATTTGCCAAGTTATATATACCAAAAGCTGTATGACTG 3218
Db 2039 TTTATGTAATCACCAGACTGGGTTATTTGCCAAGTTATATATACCAAAAGCTGTATGACTG 2098
Qy 3219 GATGCTTCTGTTACCTCGGTTTACAAATTTATCAGAGTAGT----- 3258
Db 2099 GATGCTTCTGTTACCTCGGTTTACAAATTTATCAGAGTAGTAAACTTTGATATATATGAG 2158
Qy 3259 ----- 3258
Db 2159 GATATTAATAACTACACTAAGTATCAATTTGATTTCGATTCAGAAACCTTCATATATATGAG 3278
Qy 3259 ----- 3258
Db 2159 GATATTAATAACTACACTAAGTATCAATTTGATTTCGATTCAGAAACCTTCATATATATGAG 2218


```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 162:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-09-030-607-162

Query Match          4.8%; Score 165; DB 4; Length 177;
Best Local Similarity 100.0%; Pred.No.1.4e-36;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3159 TTTATGTATCACACAGCTGGGTTATTGCCAAGTATTATATACCAAAGCTGTATGACTG 3218
      |||||||
Db 165 TTTATGTATCACACAGCTGGGTTATTGCCAAGTATTATATACCAAAGCTGTATGACTG 106
      |||||||

Qy 3219 GATGTCCTCGTTACCTGGTTTACAANAATTATCACAGTAGTAAAACITTGATATATATGAG 3278
      |||||||
Db 105 GATGTCCTCGTTACCTGGTTTACAANAATTATCACAGTAGTAAAACITTGATATATATGAG 46
      |||||||

Qy 3279 GATATAAAACACACTAGTATCATCTTTGATTCGATTCCAGAAAGT 3323
      |||||||
Db 45 GATATAAAACACACTAGTATCATCTTTGATTCGATTCCAGAAAGT 1

RESULT 6
US-09-605-785-162/c
; Sequence 162, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 162
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-605-785-162

```

```
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; NUMBER OF SEQ ID NOS: 1999-07-13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 162
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-352-616A-162

Query Match
Best Local Similarity 100.0%; Score 165; DB 4; Length 177;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3159 TTTATGTATCACCAGACTGGTTATTGCCAAGTTATATATACCAAAAGCTGTATGACTG 3218
Db 165 TTTATGTATCACCAGACTGGTTATTGCCAAGTTATATATACCAAAAGCTGTATGACTG 106

Qy 3219 GATGTTCTGGTTACCGTGGTTTACAAATATATCAGAGTAGTAAACTTTGATATATATGAG 3278
Db 105 GATGTTCTGGTTACCGTGGTTTACAAATATATCAGAGTAGTAAACTTTGATATATATGAG 46

Qy 3279 GATATTAAAACTACACTAAGTATCATTTTCGATTTCAGAAAGT 3323
Db 45 GATATTAAAACTACACTAAGTATCATTTTCGATTTCAGAAAGT 1

RESULT 9
US-09-232-149A-162/c
; Sequence 162, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 162
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-232-149A-162

Query Match
Best Local Similarity 100.0%; Score 165; DB 4; Length 177;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3159 TTTATGTATCACCAGACTGGTTATTGCCAAGTTATATATACCAAAAGCTGTATGACTG 3218
Db 165 TTTATGTATCACCAGACTGGTTATTGCCAAGTTATATATACCAAAAGCTGTATGACTG 106

Qy 3219 GATGTTCTGGTTACCGTGGTTTACAAATATATCAGAGTAGTAAACTTTGATATATATGAG 3278
Db 105 GATGTTCTGGTTACCGTGGTTTACAAATATATCAGAGTAGTAAACTTTGATATATATGAG 46

Qy 3279 GATATTAAAACTACACTAAGTATCATTTTCGATTTCAGAAAGT 3323
Db 45 GATATTAAAACTACACTAAGTATCATTTTCGATTTCAGAAAGT 1

RESULT 10
US-08-990-571-64/c
; Sequence 64, Application US/08990571
; Patent No. 6214971
```

```
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 64:
; LENGTH: 285 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-990-571-64

Query Match
Best Local Similarity 1.7%; Score 59.8; DB 4; Length 285;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 405 ATCCATATACACCATGACCGACCATCATCTACACACGAGCATCATCTGACCATCATGAG 464
Db 192 AACCATGTTCCACTAGGCCACCATCTCATCTAGGCCACCATCTCATCTAGGCCACCATG 133

Qy 465 CGTCATCAGACCATGACCATCATCTACACACGAGCATCATCTGACCATCATGATCATCAC 524
Db 132 CTTCACTAGGCCACCATCTCATCTAGGCCACCATCTCATCTAGGCCACCATCTCATCT 73

Qy 525 TCTCACCATAATCAT 539
Db 72 TAGGCCACCATGCTT 58

RESULT 11
US-09-528-784A-64/c
; Sequence 64, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 10-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Babesia microti
```


Tue Nov 26 12:43:33 2002

INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-990-571-60

Query Match 1.7%; Score 59.8; DB 4; Length 351;
Best Local Similarity 65.2%; Pred. No. 5.4e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 405 ATCCATATACACCATGACCGACGATCACTCAGACGACGATCACTCAGACCATGAG 464
DB 210 AACACGCTTCACCTAGGCCCCACGCTTCACCTAGGCCCCACGCTTCACCTAGGCCCCACGAG 151
QY 465 CGTCACTCAGACCATGACGATCACTCAGACGACGATCACTCTGACCATGATCATCAC 524
DB 150 CTTCACTAGGCCCCACGCTTCACCTAGGCCCCACGCTTCACCTAGGCCCCACGCTTCAC 91
QY 525 TCTCACCATAATCAT 539
DB 90 TAGGCCCCACGCTT 76

RESULT 15
US-09-528-784A-60/C
Sequence 60, Application US/09528784A
Patent No. 6451315
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Sleath, Paul R.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
FILE REFERENCE: 210121.426C4
CURRENT APPLICATION NUMBER: US/09/528,784A
CURRENT FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 90
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 60
LENGTH: 351
TYPE: DNA
ORGANISM: Babesia microti
US-09-528-784A-60

Query Match 1.7%; Score 59.8; DB 4; Length 351;
Best Local Similarity 65.2%; Pred. No. 5.4e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 405 ATCCATATACACCATGACCGACGATCACTCAGACGACGATCACTCAGACCATGAG 464
DB 210 AACACGCTTCACCTAGGCCCCACGCTTCACCTAGGCCCCACGCTTCACCTAGGCCCCACGAG 151
QY 465 CGTCACTCAGACCATGACGATCACTCAGACGACGATCACTCTGACCATGATCATCAC 524
DB 150 CTTCACTAGGCCCCACGCTTCACCTAGGCCCCACGCTTCACCTAGGCCCCACGCTTCAC 91
QY 525 TCTCACCATAATCAT 539
DB 90 TAGGCCCCACGCTT 76

Search completed: November 8, 2002, 20:30:56
Job time : 109.498 secs

OM nucleic - nucleic search, using sw model
Run on: November 8, 2002, 16:39:51 ; Search time 517.31 Seconds
(without alignments)
15066.735 Million cell updates/sec

Title: US-09-642-034-1
Perfect score: 3461
Sequence: 1 ctcgtgcgcaattggcagc.....aaattcttatcatcagcttg 3461

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query Match %	Length	DB ID	Description				
1	3461	100.0	3461	24	ABK92205				
2	3461	100.0	3461	24	ABA92299				
3	3391	98.0	3461	22	AAD13479				
4	3335.6	96.4	3523	23	ABV25747				
5	2761.2	79.8	2776	22	AAD13480				
6	2249	65.0	2404	19	AAT99071				
7	2245.8	64.9	2404	17	AAT33220				
8	1486.8	43.0	1597	20	AAZ33622				
9	1288	37.2	1310	19	AAT99070				

10	1284.8	37.1	1310	17	AAT33219	Oestrogen-regulate
11	1141.4	33.0	1193	21	AAF16156	Human prostate can
12	1137.8	32.9	20778	22	AAK79819	Human immune/haema
13	619.6	17.9	899	24	ABA92300	Mouse breast cancer
14	549.4	15.9	551	22	AAAD13487	Human LIV-1-164647
15	404.8	11.7	483	24	ABL79043	Human ovarian canc
16	390.8	11.3	395	22	AAH55569	Human breast tumou
17	390.4	11.3	424	22	AAH19686	Human breast cancer
18	380.4	11.0	419	22	AAH22452	Human breast cancer
19	364	10.5	366	21	AAAC30456	Human secreted pro
20	360	10.4	385	21	AAAA6009	Human metastatic m
21	340.2	9.8	603	23	ABV49440	Human prostate exp
22	319.2	9.2	337	24	ABK54054	Human head and nec
23	319	9.2	3309	23	ABV23278	Human prostate exp
24	319	9.2	3309	23	ABV29124	Human prostate exp
25	319	9.2	5322	24	ABN59760	Novel human coding
26	319	9.2	5330	22	AAF91861	Human secreted pro
27	267.6	7.7	537	23	ABV19673	Human prostate exp
28	262.4	7.6	423	22	AAH23559	Human breast cancer
29	253.8	7.3	391	22	ADL13481	Probe to detect hu
30	251.4	7.3	469	22	AAH10106	Human breast cancer
31	247	7.1	736	22	AAK92049	Human CDNA 5'-end
32	247	7.1	736	22	AAK93710	Human CDNA clone r
33	247	7.1	2863	22	AAK94381	Human full-length
34	240.8	7.0	413	22	AAH19430	Human breast cancer
35	240.8	7.0	455	22	AAH09043	Human breast cancer
36	240	6.9	241	22	AAH47302	Human breast cancer
37	240	6.9	241	22	AAH17872	Human breast cancer
38	224.8	6.5	406	23	ABV60690	Human prostate exp
39	206.6	6.0	643	22	AAH18330	Human breast cancer
40	189.4	5.5	405	22	AAH18310	Human breast cancer
41	165	4.8	177	19	AAV61316	cDNA sequence of p
42	165	4.8	177	19	AAV58635	Prostate tumour sp
43	165	4.8	177	21	AAH06398	Human immunogenic
44	165	4.8	177	22	AAH63606	Human prostate cDN
45	165	4.8	177	22	AAH10157	Human prostate tum

ALIGNMENTS

RESULT 1	
ABK92205	ABK92205 standard; DNA; 3461 BP.
ID	ABK92205 standard; DNA; 3461 BP.
AC	ABK92205;
DT	15-AUG-2002 (first entry)
DE	Prostate cancer-associated DNA sequence #91.
KW	Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
KW	gene therapy; gene; ds.
OS	Mammalia.
PN	WO200230268-A2.
XX	18-APR-2002.
PD	12-OCT-2001; 2001WO-US32045.
PF	13-DEC-2000; 2000US-0687576.
PR	08-DEC-2000; 2000US-0733288.
PR	08-DEC-2000; 2000US-0733742.
PR	24-JAN-2001; 2001US-263957P.
PR	16-MAR-2001; 2001US-276791P.
PR	06-APR-2001; 2001US-276888P.
PR	24-APR-2001; 2001US-281922P.
PR	30-APR-2001; 2001US-286214P.
PR	04-MAY-2001; 2001US-0847046.
XX	

PA (EOSB-) EOS BIOTECHNOLOGY INC.
PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX
XX
XX WPI: 2002-471335/50.
DR P-PSDB: ABG61889.
XX
XX Detecting a prostate cancer-associated transcript in a cell in a
PT patient, useful for diagnosing prostate cancer (PC) or screening
PT modulators of PC, by determining if prostate cancer-associated genes
PT are expressed in a prostate tissue
XX
XX Claim 22; Page 374-375; 436pp; English.
XX
XX The present invention relates to methods of detecting a prostate
CC cancer-associated transcript in a cell from a patient. The method
CC comprises contacting a biological sample from the patient with
CC prostate cancer-associated polynucleotides (designated PC genes) that
CC selectively hybridize to a sequence that is at least 80% identical
CC to them. The prostate cancer-associated polynucleotide sequences
CC are differentially expressed in prostate tumour tissue or in
CC prostate cancer and are derived from the tissues of various
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC The methods of the invention are useful for diagnosing and treating
CC prostate cancer in mammals. The prostate cancer-associated genes are
CC useful for diagnosing or treating prostate cancer, as well as for
CC identifying modulators of prostate cancer or agents that inhibit
CC prostate cancer. The nucleic acid sequences are particularly useful
CC in gene therapy, as a vaccine or in antisense applications.
CC ASK92115-ABK92263 represent prostate cancer-associated polynucleotide
CC sequences.
XX
XX Sequence 3461 BP; 1042 A; 686 C; 728 G; 1005 T; 0 other;
S0
Query Match 100.0%; Score 3461; DB 24; Length 3461;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTCGTCGCGAATTCGGCAGCAGACCGCTGTCGCGCTGGTAGAGATTCTCGAAGACA 60
Db 1 CTCGTCGCGAATTCGGCAGCAGACCGCTGTCGCGCTGGTAGAGATTCTCGAAGACA 60
Qy 61 CCAGTGGGCGCGTGTGGAACAAACCTGCGCGCTGGCGGGCGGTGGGAACACGAGGCC 120
Db 61 CCAGTGGGCGCGTGTGGAACAAACCTGCGCGCTGGCGGGCGGTGGGAACACGAGGCC 120
Qy 121 GCGGAGCAGAGGCGCAATGCGGAGGAGTATCTGTAATCTTGATCCTTGACCTTTGCC 180
Db 121 GCGGAGCAGAGGCGCAATGCGGAGGAGTATCTGTAATCTTGATCCTTGACCTTTGCC 180
Qy 181 TCTCTGTACAAATCCCTTCATGAACATAAAGCAGCTGCTTCCCGCCAGACCACTGAGA 240
Db 181 TCTCTGTACAAATCCCTTCATGAACATAAAGCAGCTGCTTCCCGCCAGACCACTGAGA 240
Qy 241 AAATTAGTCGGAATTCGGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAAT 300
Db 241 AAATTAGTCGGAATTCGGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAAT 300
Qy 301 ATCATCTACACAGCTTTCTACCGCTATGGAGAAATAATCTTCTCAGTTGAGGGT 360
Db 301 ATCATCTACACAGCTTTCTACCGCTATGGAGAAATAATCTTCTCAGTTGAGGGT 360
Qy 361 TCAGAAATTAATCTCAAAATATAGGCATAGATAGATTAAGAAATCCATATACACCATG 420
Db 361 TCAGAAATTAATCTCAAAATATAGGCATAGATAGATTAAGAAATCCATATACACCATG 420
Qy 421 ACCAGGACCATCTCAGACACAGCATCTCAGACCATGAGCGTCACCTCAGACCATG 480
Db 421 ACCAGGACCATCTCAGACACAGCATCTCAGACCATGAGCGTCACCTCAGACCATG 480
Qy 481 AGCATCTCAGACACGAGCATCTCTGACCATGATCATCTCTCACCATAATCATG 540
Db 481 AGCATCTCAGACACGAGCATCTCTGACCATGATCATCTCTCACCATAATCATG 540

Qy 541 CTGCTTCTGGTAAATAAGCGAAAGCTCTTTGCCAGACCATGACTCAGATGATTTCAG 600
Db 541 CTGCTTCTGGTAAATAAGCGAAAGCTCTTTGCCAGACCATGACTCAGATGATTTCAG 600
Qy 601 GTAAAGATCTCTAGAAACAGCCAGCGGAAAGAGCTCACCGACCAAGACATGCCAGTGGTA 660
Db 601 GTAAAGATCTCTAGAAACAGCCAGCGGAAAGAGCTCACCGACCAAGACATGCCAGTGGTA 660
Qy 661 GAAGAAATGTCAGGACAGCTGTTAGTCTAGTGAAGTGACCTCAACTGTGTACAACTG 720
Db 661 GAAGAAATGTCAGGACAGCTGTTAGTCTAGTGAAGTGACCTCAACTGTGTACAACTG 720
Qy 721 TCTCTGAAGAACTCAGCTTTCTAGAGACATAGAGACTTCCAAAGACCTTGGAAACCTCTTCC 780
Db 721 TCTCTGAAGAACTCAGCTTTCTAGAGACATAGAGACTTCCAAAGACCTTGGAAACCTCTTCC 780
Qy 781 CCAAAGATGTAAGCAGCTTCCACTCCACCCAGTGTACATCAAGAGCGGGGTGAGCGGC 840
Db 781 CCAAAGATGTAAGCAGCTTCCACTCCACCCAGTGTACATCAAGAGCGGGGTGAGCGGC 840
Qy 841 TGGCTGGTAGGAAACAAATGAATCTGTGAGTGAGCCCCGAAAGGCTTTATGTTATCCA 900
Db 841 TGGCTGGTAGGAAACAAATGAATCTGTGAGTGAGCCCCGAAAGGCTTTATGTTATCCA 900
Qy 901 GAAACACAAATGAAATCCTCAGGAGTGTTCATGATCAAAAGCTACTGACATCTCATG 960
Db 901 GAAACACAAATGAAATCCTCAGGAGTGTTCATGATCAAAAGCTACTGACATCTCATG 960
Qy 961 GCATGGGATCCAGTTCGGTGAATGCAACAGAGTTCAACTATCTGTGTCCAGCCATCA 1020
Db 961 GCATGGGATCCAGTTCGGTGAATGCAACAGAGTTCAACTATCTGTGTCCAGCCATCA 1020
Qy 1021 TCAACCAAAATGATGCTAGATCTTGTCTGATTATACAAAGTGAAGAGCTGAAATCC 1080
Db 1021 TCAACCAAAATGATGCTAGATCTTGTCTGATTATACAAAGTGAAGAGCTGAAATCC 1080
Qy 1081 CTCCTAAAGACCTTATTCATTACAAATAGCTGGTGGTGGTATATAGCCATTTCCATCA 1140
Db 1081 CTCCTAAAGACCTTATTCATTACAAATAGCTGGTGGTGGTATATAGCCATTTCCATCA 1140
Qy 1141 TCAGTTTCTCTGCTCTGCTGGGGGTATCTTAGTGCCTCTCATGAATCGGCTGTTTTCA 1200
Db 1141 TCAGTTTCTCTGCTCTGCTGGGGGTATCTTAGTGCCTCTCATGAATCGGCTGTTTTCA 1200
Qy 1201 AATTTCTCTGAGTTTCTTGTGGCAGCTGGCGTTGGGACTTTGAGTGGTGTATGCTTTT 1260
Db 1201 AATTTCTCTGAGTTTCTTGTGGCAGCTGGCGTTGGGACTTTGAGTGGTGTATGCTTTT 1260
Qy 1261 TACACCTTCTCCACATTCATGCAAGTCAACCATAGTCAAGCCATAGCCATGAAGAACCCAG 1320
Db 1261 TACACCTTCTCCACATTCATGCAAGTCAACCATAGTCAAGCCATAGCCATGAAGAACCCAG 1320
Qy 1321 CAATGGAATGAAAGAGAGCCACCTTTTCAGTCATCTGCTCTCTCAAAACATAGAAGAAA 1380
Db 1321 CAATGGAATGAAAGAGAGCCACCTTTTCAGTCATCTGCTCTCTCAAAACATAGAAGAAA 1380
Qy 1381 GTGCTTATTTGATTCACAGTGGAAAGGCTTAACAGCTCTTAGGAGCCCTGTATTTTCATGT 1440
Db 1381 GTGCTTATTTGATTCACAGTGGAAAGGCTTAACAGCTCTTAGGAGCCCTGTATTTTCATGT 1440
Qy 1441 TTCTTGTGAAACATGCTCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1441 TTCTTGTGAAACATGCTCTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Qy 1501 AGAAGAAACCTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1501 AGAAGAAACCTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Qy 1561 CTCACCTTTCAACAAATGAGGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1620
Db 1561 CTCACCTTTCAACAAATGAGGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1620

Qy	1621	GAGCAGACTCACAGAGCCCTCCCACTTTTGATTTCTCAGAGCGCTGCAGTCTTTGGAAGAAG	1681
Db	1621	GAGCAGACTCACAGAGCCCTCCCACTTTTGATTTCTCAGAGCGCTGCAGTCTTTGGAAGAAG	1680
Qy	1681	AAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTATCCCAAG	1740
Db	1681	AAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTATCCCAAG	1740
Qy	1741	GGTGCAAGAAATAATGCCATTACATTTCCACGATACACTCGGCCAGTCACAGCATCTCTCA	1800
Db	1741	GGTGCAAGAAATAATGCCATTACATTTCCACGATACACTCGGCCAGTCACAGCATCTCTCA	1800
Qy	1801	TTTACCACCATTATGACTACCATCATPATTTCTCCATCATCACACACACCAAAACCACCATC	1860
Db	1801	TTTACCACCATTATGACTACCATCATPATTTCTCCATCATCACACACCAAAACCACCATC	1860
Qy	1861	CTCAGCTCACGCCAGCGCTACTCTCGGGAGGAGCTGAAAGATGCGCGGCTCGCCACTT	1920
Db	1861	CTCAGCTCACGCCAGCGCTACTCTCGGGAGGAGCTGAAAGATGCGCGGCTCGCCACTT	1920
Qy	1921	TGGCCTGGATGGTGATAATGGGTGATGGCGTCGCACAAATTTAGCGCATGGCCCTAGCAATTG	1980
Db	1921	TGGCCTGGATGGTGATAATGGGTGATGGCGTCGCACAAATTTAGCGCATGGCCCTAGCAATTG	1980
Qy	1981	GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTTAAGTACTTCTGTTGCTGTCTGTCTGTC	2040
Db	1981	GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTTAAGTACTTCTGTTGCTGTCTGTCTGTC	2040
Qy	2041	ATGAGTTGGCCTCATGAATTAGTGACTTTGCTTGTCTTACTTAAGGCTGGCATGACCGTTA	2100
Db	2041	ATGAGTTGGCCTCATGAATTAGTGACTTTGCTTGTCTTACTTAAGGCTGGCATGACCGTTA	2100
Qy	2101	AGCAGGCTGTCCCTTATATATGCAATGTCAAGCCATGCTGGCGTATCTTGGAAATGGCAACAG	2160
Db	2101	AGCAGGCTGTCCCTTATATATGCAATGTCAAGCCATGCTGGCGTATCTTGGAAATGGCAACAG	2160
Qy	2161	GAATTTTTCATTTGCTCATTTATGCTGAAAAATGTTTCTATGTGGATATTTGCACTTACTGCTG	2220
Db	2161	GAATTTTTCATTTGCTCATTTATGCTGAAAAATGTTTCTATGTGGATATTTGCACTTACTGCTG	2220
Qy	2221	GCATTATTCATGTATGTTGCTCTGGTTGATATGGTACCTGAAATGCTGCACAAATGATGCTA	2280
Db	2221	GCATTATTCATGTATGTTGCTCTGGTTGATATGGTACCTGAAATGCTGCACAAATGATGCTA	2280
Qy	2281	GTGACCATTGATGTAGCCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG	2340
Db	2281	GTGACCATTGATGTAGCCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG	2340
Qy	2341	GTTTTTGGAAATTTGTTACTTTATTTCCAAATTTGAACATATAATCGTGTATATAAAT	2400
Db	2341	GTTTTTGGAAATTTGTTACTTTATTTCCAAATTTGAACATATAATCGTGTATATAAAT	2400
Qy	2401	TCTAGTTTAAAGGTTTAAATGCTAGGTAGCTTAAAAGTTTGTACATAGTTTTCAGTAGGTCAT	2460
Db	2401	TCTAGTTTAAAGGTTTAAATGCTAGGTAGCTTAAAAGTTTGTACATAGTTTTCAGTAGGTCAT	2460
Qy	2461	AGGGAGATGAGTTTGTATGCTGTACTATGCACGCGTTTAAAAGTTAGTGCGTTTCTGATTT	2520
Db	2461	AGGGAGATGAGTTTGTATGCTGTACTATGCACGCGTTTAAAAGTTAGTGCGTTTCTGATTT	2520
Qy	2521	TTGTATTTGAAATTTGCTGTCTGTTTACAAAGTCAGTTAAAGSTACGTTTTAATTTAAGT	2580
Db	2521	TTGTATTTGAAATTTGCTGTCTGTTTACAAAGTCAGTTAAAGSTACGTTTTAATTTAAGT	2580
Qy	2581	TATCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAGTTTTATTATG	2640
Db	2581	TATCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAGTTTTATTATG	2640
Qy	2641	TAAACAAGAGATTTGGCATCAGATGTTCTGTATGTTTACAGGAAAAAATGCTCTTTAATGCT	2700
Db	2641	TAAACAAGAGATTTGGCATCAGATGTTCTGTATGTTTACAGGAAAAAATGCTCTTTAATGCT	2700
Qy	2701	TTTTTCAAGAACTAACACAGTTATTTCTTATACTGGATTTTTAGTCTCTCGAAGAACTGCTGG	2760

[illegible]

Tue Nov 26 12:43:33 2002

FT mat_peptide 198..2402
FT /*tag- c
XX W0200216939-A2.
XX 28-FEB-2002.
XX PD
XX PF
XX PP
XX 20-AUG-2001; 2001WO-US25997.
XX 18-AUG-2000; 2000US-0642034.
XX PR 08-DEC-2000; 2000US-0733320.
XX PA (BOSB-) EOS BIOTECHNOLOGY INC.
XX PI Mack D, Glish KC, Wilson KE;
XX WPI; 2002-242033/29.
XX P-PSDB; AAM51198.

Screening for drugs that affect expression of the breast cancer 4 gene or its fragments, use of these to treat prostate and breast cancer, and diagnosing these diseases -
Claim 27; Fig 1A-B; 83pp; English.

The presence sequence is that of cDNA identified for the human breast cancer 4 gene (BCR4) on chromosome 18 (cytoband 18q12). The gene was identified as being up-regulated in specific cancer types in an oligonucleotide microarray interrogated with cRNAs derived from multiple tissues. The gene encodes the protein given in AAM51198. The invention provides claimed methods of using BCR4 polynucleotides, polypeptides and antibodies to: screen drug candidates; screen bioactive agents capable of binding to BCR4; evaluate the effect of a prostate cancer and/or breast cancer drug; diagnose prostate cancer or breast cancer; screen for a bioactive agent capable of interfering with the binding of BCR4 to an antibody which binds BCR4; neutralise the effect of BCR4; treat prostate cancer or breast cancer using an inhibitor of BCR4 (e.g. an antibody); localise a therapeutic moiety to, or treat, prostate cancer or breast cancer tissue by conjugating an antibody for BCR4 to a cytotoxic agent or radioisotope; inhibit prostate cancer or breast cancer by administering an antisense molecule to the present nucleic acid sequence; elicit an immune response by administering BCR4 or a nucleic acid encoding BCR4; and determine the prognosis of an individual with prostate cancer or breast cancer by determining the level of BCR4 in a sample, a high level indicating a poor prognosis. A biochip comprising BCR4 nucleic acids is also claimed.

Sequence 3461 BP; 1042 A; 686 C; 728 G; 1005 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 3461; DB 24; Length 3461;
Matches 3461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCGTGCCGAATTCGGCAGACGCGGTTCGCGCGCTGGTAGAGATTTCTCGAAGACA 60
DB 1 CTCGTGCCGAATTCGGCAGACGCGGTTCGCGCGCTGGTAGAGATTTCTCGAAGACA 60
QY 61 CCAGTGGCCGCTGTGGAAACCAACCTCGCGCGCTGGCGCGCGCTGGGACACGAGGCC 120
DB 61 CCAGTGGCCGCTGTGGAAACCAACCTCGCGCGCTGGCGCGCGCTGGGACACGAGGCC 120
QY 121 GCGGAGACGAAGCGCAATGGCGAGGAAGTTATCTGTATCTTGATCTGACCTTGCCGCC 180
DB 121 GCGGAGACGAAGCGCAATGGCGAGGAAGTTATCTGTATCTTGATCTGACCTTGCCGCC 180
QY 181 TCTCTGTCAAAATCCCTTCATGAACATAAAGCAGCTGCTTCCCGCCAGACCACTCGAGA 240
DB 181 TCTCTGTCAAAATCCCTTCATGAACATAAAGCAGCTGCTTCCCGCCAGACCACTCGAGA 240
QY 241 AAATTAGTCCGAATTCGGGAATCTGGCAATTAATGTTGACTTGGCAATTTCCACACGGCAAT 300
DB 241 AAATTAGTCCGAATTCGGGAATCTGGCAATTAATGTTGACTTGGCAATTTCCACACGGCAAT 300

DB 241 AAATTAGTCCGAATTCGGGAATCTGGCAATTAATGTTGACTTGGCAATTTCCACACGGCAAT 300
QY 301 ATCATCTACAACAGCTTTTCTACCGCTATGGAGAAAATAATCTTTGTCTAGTTGAAGGGT 360
DB 301 ATCATCTACAACAGCTTTTCTACCGCTATGGAGAAAATAATCTTTGTCTAGTTGAAGGGT 360
QY 361 TCAGAAAATTTACTTCAAAATATAGCATAGATAAGATTAAAGAAATCCATATACACCATG 420
DB 361 TCAGAAAATTTACTTCAAAATATAGCATAGATAAGATTAAAGAAATCCATATACACCATG 420
QY 421 ACCCAGCACCATCACTCAGACCCAGCAGCATCTCAGACCATGAGCGTCACTCAGACCATG 480
DB 421 ACCCAGCACCATCACTCAGACCCAGCAGCATCTCAGACCATGAGCGTCACTCAGACCATG 480
QY 481 AGCATCACTCAGACCCAGCAGCATCTCAGACCATGAGCGTCACTCAGACCATGAGCGTCA 540
DB 481 AGCATCACTCAGACCCAGCAGCATCTCAGACCATGAGCGTCACTCAGACCATGAGCGTCA 540
QY 541 CTGCTTCTGGTAAAAATAAGCGAAAGCTTTTGGCCAGACCATGACTCAGATAGTTTCAG 600
DB 541 CTGCTTCTGGTAAAAATAAGCGAAAGCTTTTGGCCAGACCATGACTCAGATAGTTTCAG 600
QY 601 GTAAAGATCCTAGAACACGCGAGGGAAGAGCTCACCGACGACCAACATGCCAGTGTGA 660
DB 601 GTAAAGATCCTAGAACACGCGAGGGAAGAGCTCACCGACGACCAACATGCCAGTGTGA 660
QY 661 GAAGGAATGTCAAGGACAGTGTGTAGTGTAGTGAAGTGAACCTCAACTGTGTACAACTG 720
DB 661 GAAGGAATGTCAAGGACAGTGTGTAGTGTAGTGAAGTGAACCTCAACTGTGTACAACTG 720
QY 721 TCTCTGAAGAACTCACTTTCTAGACACAATAGAGACTCCAAGACTGGGAAACTCTTCC 780
DB 721 TCTCTGAAGAACTCACTTTCTAGACACAATAGAGACTCCAAGACTGGGAAACTCTTCC 780
QY 781 CCAAGATGTAAAGCAGCTCCACCTCCAGCTGCACATCAAGAGCGCGGTGAGCCGGC 840
DB 781 CCAAGATGTAAAGCAGCTCCACCTCCAGCTGCACATCAAGAGCGCGGTGAGCCGGC 840
QY 841 TGGCTGTAGGAAACAAATGAATCTGTAGTGAAGCTCCGCGGAAAGGCTTTATGTATCCA 900
DB 841 TGGCTGTAGGAAACAAATGAATCTGTAGTGAAGCTCCGCGGAAAGGCTTTATGTATCCA 900
QY 901 GAAACACAATGAAATTCCTCAGAGTGTTCATATGCATCAAGCTCACTGACATCTCATG 960
DB 901 GAAACACAATGAAATTCCTCAGAGTGTTCATATGCATCAAGCTCACTGACATCTCATG 960
QY 961 GCATGGGCATCCAGGTCGCGTGAATGCAACAGATTCACACTATCTCTGTCAGCCATCA 1020
DB 961 GCATGGGCATCCAGGTCGCGTGAATGCAACAGATTCACACTATCTCTGTCAGCCATCA 1020
QY 1021 TCAACCAATTTGATGCTAGATCTTGTCTGATTCATACAAGTGAAAGAGGCTGAAATCC 1080
DB 1021 TCAACCAATTTGATGCTAGATCTTGTCTGATTCATACAAGTGAAAGAGGCTGAAATCC 1080
QY 1081 CTCCAAAGACCTATTTCATTAACAAATAGCCTGGGTTGGTGGTTTATAGCCATTTCCATCA 1140
DB 1081 CTCCAAAGACCTATTTCATTAACAAATAGCCTGGGTTGGTGGTTTATAGCCATTTCCATCA 1140
QY 1141 TCAGTTTCTCTGCTGCTGCTGGGGTTATCTTAGTGCCTCTCATGAATCGGGTGTTTTCA 1200
DB 1141 TCAGTTTCTCTGCTGCTGCTGGGGTTATCTTAGTGCCTCTCATGAATCGGGTGTTTTCA 1200
QY 1201 AATTTCTCTGAGTTTCTTGTGGCAGTGGCGGTGGGACTTTGAGTGGTGTGCTTTT 1260
DB 1201 AATTTCTCTGAGTTTCTTGTGGCAGTGGCGGTGGGACTTTGAGTGGTGTGCTTTT 1260
QY 1261 TACACCTTCTCCACATTTCTCATGCAAGTCAACCATAGTCAATAGCCATGGAAGACGAG 1320
DB 1261 TACACCTTCTCCACATTTCTCATGCAAGTCAACCATAGTCAATAGCCATGGAAGACGAG 1320
QY 1321 CAATGGAATGAAAGAGGAGGACCACTTTTTCAGTCACTCTCTCTCAAAACATAGAAGAA 1380
DB 1321 CAATGGAATGAAAGAGGAGGACCACTTTTTCAGTCACTCTCTCTCAAAACATAGAAGAA 1380

|||||
Db 1243 TACACCTCTTCCACATCTCATGCAAGTCACCCACCATAGTCATAGCCATGAAGAACCAG 1302
|||||
Qy 1321 CAATGGAATGAAAGAGGAGGACCACTTTTCAGTCATCTGTCTCTCAAAACATAGAGAAA 1380
|||||
Db 1303 CAATGGAATGAAAGAGGAGGACCACTTTTCAGTCATCTGTCTCTCAAAACATAGAGAAA 1362
|||||
Qy 1381 GTGCCATTTTTCATTTCCACGCTGGAAGGGCTTAACAGCTCTAGGAGCCCTGTATTTTCATGT 1440
|||||
Db 1363 GTGCCATTTTTCATTTCCAGCTGGAAGGGCTTAACAGCTCTAGGAGCCCTGTATTTTCATGT 1422
|||||
Qy 1441 TTCTTGTTCACACATGTCTTCACATTCATGATCAAAACAAATTAAGAGATGAAGAGAAAAGATC 1500
|||||
Db 1423 TTCTTGTTCACACATGTCTTCACATTCATGATCAAAACAAATTAAGAGATGAAGAGAAAAGATC 1482
|||||
Qy 1501 AGAAGAAACCTGAAATGATGATGTGGAGATTAAAGAGCAGTTGTCCAGTATGAAT 1560
|||||
Db 1483 AGAAGAAACCTGAAATGATGATGTGGAGATTAAAGAGCAGTTGTCCAGTATGAAT 1542
|||||
Qy 1561 CTCAACTTTCAACAAATGAGGAGAAAGTAGATACAGATGATCGAACTGAAGGCTATTTCAC 1620
|||||
Db 1543 CTCAACTTTCAACAAATGAGGAGAAAGTAGATACAGATGATCGAACTGAAGGCTATTTCAC 1602
|||||
Qy 1621 GAGCAGCTCACAAAGAGCCCTCCACTTTTGATTTCTCAGCAGCTCGAGTCTTTGGAGAGAG 1680
|||||
Db 1603 GAGCAGCTCACAAAGAGCCCTCCACTTTTGATTTCTCAGCAGCTCGAGTCTTTGGAGAGAG 1662
|||||
Qy 1681 ACAGGTCATGATGATGATGTCTCCTCACAGGAGTCTACAAATGAATGTATACCCAGAG 1740
|||||
Db 1663 ACAGGTCATGATGATGATGTCTCCTCACAGGAGTCTACAAATGAATGTATACCCAGAG 1722
|||||
Qy 1741 GGTGCAAGAAATAATGCCATTCACATTTCCACGATACACTCGGCCAGTCAGACGATCTCA 1800
|||||
Db 1723 GGTGCAAGAAATAATGCCATTCACATTTCCACGATACACTCGGCCAGTCAGACGATCTCA 1782
|||||
Qy 1801 TTCACACCATCATGACTACCATCATATTTCTCCATCATCACCACCAACCAACCCACCATC 1860
|||||
Db 1783 TTCACACCATCATGACTACCATCATATTTCTCCATCATCACCACCAACCAACCCACCATC 1842
|||||
Qy 1861 CTCACAGTCACAGCCAGCGCTACTCTCGGGAGGAGCTGAAAGATCGCGCGCTCGGCACATT 1920
|||||
Db 1843 CTCACAGTCACAGCCAGCGCTACTCTCGGGAGGAGCTGAAAGATCGCGCGCTCGGCACATT 1902
|||||
Qy 1921 TGGCTGGATGGATGAATGGGTGATGGCTGCACAAATTTACAGGATGGCCCTAGCAATTCG 1980
|||||
Db 1903 TGGCTGGATGGATGAATGGGTGATGGCTGCACAAATTTACAGGATGGCCCTAGCAATTCG 1962
|||||
Qy 1981 GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAGTACTTCTGTGTGCTGTCTGTCTC 2040
|||||
Db 1963 GTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAGTACTTCTGTGTGCTGTCTGTCTC 2022
|||||
Qy 2041 ATGAGTTGCCATCATGAATAGGTGACTTTTGTCTTCTACTAAAGGCTGGCATGACCGTTA 2100
|||||
Db 2023 ATGAGTTGCCATCATGAATAGGTGACTTTTGTCTTCTACTAAAGGCTGGCATGACCGTTA 2082
|||||
Qy 2101 AGCAGGCTGTCTTTTATATGCAATTTGTACCCCATGCTGGCGTATCTTGGATGGCAACAG 2160
|||||
Db 2083 AGCAGGCTGTCTTTTATATGCAATTTGTACCCCATGCTGGCGTATCTTGGATGGCAACAG 2142
|||||
Qy 2161 GAATTTTTCATTTGTCATTTGCTGAAATGCTTTCTATGTGGATTTTGTGACTTTACTGCTG 2220
|||||
Db 2143 GAATTTTTCATTTGTCATTTGCTGAAATGCTTTCTATGTGGATTTTGTGACTTTACTGCTG 2202
|||||
Qy 2221 GCTTATTCATGTATGTGCTCTGTTGATATGGTACCTGAAATGCTGCACAAATGATGCTA 2280
|||||
Db 2203 GCTTATTCATGTATGTGCTCTGTTGATATGGTACCTGAAATGCTGCACAAATGATGCTA 2262
|||||
Qy 2281 GTGACCATGGATGAGCCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG 2340
|||||
Db 2263 GTGACCATGGATGAGCCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG 2322
|||||
Qy 2341 GTTTTGGAAATTTACTTATTTTCCATATTTGAACATAAAATCGTGTCTTTCGTATAAAT 2400
|||||

Db 2323 GTTTTGGAAATTTACTTTA - TTCCATATTTGAACATAAAATCGTG - TTCGTATAAAAT 2380
Qy 2401 TCTAGTTAAAGTTTAAATGCTAGAGTAGCTTAAAGCTTGTCTAGTTTTCAGTAGGTCAT 2460
|||||
Db 2381 TCTAGTTAAAGTTTAAATGCTAGAGTAGCTTAAAGCTTGTCTAGTTTTCAGTAGGTCAT 2440
|||||
Qy 2461 AGGAGATGAGTTTGTATGCTGTACTATGTCAGCGCTTTAAAGTTAGTGGGTTTTGTGATTT 2520
|||||
Db 2441 AGGAGATGAGTTTGTATGCTGTACTATGTCAGCGCTTTAAAGTTAGTGGGTTTTGTGATTT 2500
|||||
Qy 2521 TTGTATTGAATATTTGCTGTCTGTCTACAAAGTCAGTTAAAGTCAGTTTAAATATTAAAT 2580
|||||
Db 2501 TTGTATTGAATATTTGCTGTCTGTCTACAAAGTCAGTTAAAGTCAGTTTAAATATTAAAT 2560
|||||
Qy 2581 TATTTCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAAGTTTATTATG 2640
|||||
Db 2561 TATTTCTATCTTGGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAAGTTTATTATG 2620
|||||
Qy 2641 TAAACAAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGAGAAAATGCTTTTAAATGCT 2700
|||||
Db 2621 TAAACAAGAGATTTGGCATGACATGTTCTGTATGTTTCAGGAGAAAATGCTTTTAAATGCT 2680
|||||
Qy 2701 TTTTCAAGAACTAACACAGTTTATTCCTATACTGGAATTTTAGGTCCTCTGAAGAACTGCTGG 2760
|||||
Db 2681 TTTTCAAGAACTAACACAGTTTATTCCTATACTGGAATTTTAGGTCCTCTGAAGAACTGCTGG 2740
|||||
Qy 2761 TGTTTAGGAATAAGAACTGTGCATGAAGCCTAAATACCAAGAACTTATAGGTCCTCTGAAGAACTGCTGG 2820
|||||
Db 2741 TGTTTAGGAATAAGAACTGTGCATGAAGCCTAAATACCAAGAACTTATAGGTCCTCTGAAGAACTGCTGG 2800
|||||
Qy 2821 AGCAAGAAATTAAGGAGAAAGAGAGAACTCTGAGAAATTTGGGAGGAGCATAGATCTTTAT 2880
|||||
Db 2801 AGCAAGAAATTAAGGAGAAAGAGAGAACTCTGAGAAATTTGGGAGGAGCATAGATCTTTAT 2860
|||||
Qy 2881 AAAATTCACAAATTTGTTGTAATTTAGAGGGAGAAATTTAGAAATTAAGTATATAAAGG 2940
|||||
Db 2861 AAAATTCACAAATTTGTTGTAATTTAGAGGGAGAAATTTAGAAATTAAGTATATAAAGG 2920
|||||
Qy 2941 CAGAAATAGTAGAGTACATTTCAATTAACATTTTGTGAGGATTTTCCCGTAAAAAC 3000
|||||
Db 2921 CAGAAATAGTAGAGTACATTTCAATTAACATTTTGTGAGGATTTTCCCGTAAAAAC 2980
|||||
Qy 3001 GTAGTGAGCAGCTCATATACTAATTTAGTGTACATTTTAACTTTGTATATAACAGAAATCT 3060
|||||
Db 2981 GTAGTGAGCAGCTCATATACTAATTTAGTGTACATTTTAACTTTGTATATAACAGAAATCT 3040
|||||
Qy 3061 AAATATATTTAATGAATTCAGCAATATACACTTGACCAAGAAATTTGAAATTTCAAAATG 3120
|||||
Db 3041 AAATATATTTAATGAATTCAGCAATATACACTTGACCAAGAAATTTGAAATTTCAAAATG 3100
|||||
Qy 3121 TTCGTGCGGGTTATATACAGATGAGTACAGTGTAGTGTATTTATGTATCACCAGACTGGGT 3180
|||||
Db 3101 TTCGTGCGGGTTATATACAGATGAGTACAGTGTAGTGTATTTATGTATCACCAGACTGGGT 3160
|||||
Qy 3181 TATTTCCCAAGTTATATATACCAAAAGCTGTATGACTGATGCTGCTGCTGCTGCTGCTGCTG 3240
|||||
Db 3161 TATTTCCCAAGTTATATATACCAAAAGCTGTATGACTGATGCTGCTGCTGCTGCTGCTGCTG 3220
|||||
Qy 3241 CAAAATATTCAGAGTAGTAAATCTTTGATATATATAGGATTTTAAAACTACACTAAGTA 3300
|||||
Db 3221 CAAAATATTCAGAGTAGTAAATCTTTGATATATATAGGATTTTAAAACTACACTAAGTA 3280
|||||
Qy 3301 TCATTTGATTCAGATTCAGAAAGTACTTTGATATCTCTCAGTGTCTCAGTGTCTCATTTGT 3360
|||||
Db 3281 TCATTTGATTCAGATTCAGAAAGTACTTTGATATCTCTCAGTGTCTCAGTGTCTCATTTGT 3340
|||||
Qy 3361 GAGCAATTCCTTTTATATACGCTACTGTAGCCATACTAGGCTGCTGCTGGCATTTCTCTA 3420
|||||
Db 3341 GAGCAATTCCTTTTATATACGCTACTGTAGCCATACTAGGCTGCTGCTGGCATTTCTCTA 3400
|||||
Qy 3421 GATGTTCTCTTTTACACAATAAAATTCCTTATATCAGCTTG 3461
|||||
Db 3401 GATGTTCTCTTTTACACAATAAAATTCCTTATATCAGCTTG 3441
|||||

Db 1185 TTCTCCTGAGTTCTCTGGCACTGGCCGTTGGGACTTTGAGTGGTATGCTTTTTTA 1244
Qy 1263 CACCTTCTCCACATCTCATCGAAGTCACCACTAGTCATAGCATGAAGAACGACA 1322
Db 1245 CACCTTCTCCACATCTCATCGAAGTCACCACTAGTCATAGCATGAAGAACGACA 1304
Qy 1323 ATGGAATGAAGAGGACCACTTTTCAGTCATCTGCTCTCTCAAAACATGAAGAAAGT 1382
Db 1305 ATGGAATGAAGAGGACCACTTTTCAGTCATCTGCTCTCTCAAAACATGAAGAAAGT 1364
Qy 1383 GCCTATTTTCATTCACAGCTGGAAGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTT 1442
Db 1365 GCCTATTTTCATTCACAGCTGGAAGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTT 1424
Qy 1443 CTTGTTGAACATGCTCCTCAGTTCATCAACCAATTTTAAAGATGAAGAAAGAAATCAG 1502
Db 1425 CTTGTTGAACATGCTCCTCAGTTCATCAACCAATTTTAAAGATGAAGAAAGAAATCAG 1484
Qy 1503 AGAAACCTGAAATGATGATGTGGAGATTGAAGCAGTTGTCCAAAGTATGATCT 1562
Db 1485 AGAAACCTGAAATGATGATGTGGAGATTGAAGCAGTTGTCCAAAGTATGATCT 1544
Qy 1563 CAACCTTCAACAAATGAGGAAAGTAGATACAGATGATCGAAGTCAAGGCTATTTACGA 1622
Db 1545 CAACCTTCAACAAATGAGGAAAGTAGATACAGATGATCGAAGTCAAGGCTATTTACGA 1604
Qy 1623 GCAGACTCACAAGAGCCCTCCCACTTTGATTTCTCAGCAGCCTGCGAGTGTGAAGAA 1682
Db 1605 GCAGACTCACAAGAGCCCTCCCACTTTGATTTCTCAGCAGCCTGCGAGTGTGAAGAA 1664
Qy 1683 GAGTCAATGATGCTCATGCTCATCCACAGGAGTCTACAAATGATATGACCCAGAGG 1742
Db 1665 GAGTCAATGATGCTCATGCTCATCCACAGGAGTCTACAAATGATATGACCCAGAGG 1724
Qy 1743 TGCAAGAAATAAATGCCATTCACATTTCCACGATACACTCGGCCAGTCAGCAGATCTCAT 1802
Db 1725 TGCAAGAAATAAATGCCATTCACATTTCCACGATACACTCGGCCAGTCAGCAGATCTCAT 1784
Qy 1803 CACCACCATCATGATACCATCATATTTCTCCATCATCACCACCAACCAACCATCT 1862
Db 1785 CACCACCATCATGATACCATCATATTTCTCCATCATCACCACCAACCAACCATCT 1844
Qy 1863 CACAGTCACAGCCAGCCCTACTCTCGGGAGAGCTGAAGATGCCGCGTCGCCACTTG 1922
Db 1845 CACAGTCACAGCCAGCCCTACTCTCGGGAGAGCTGAAGATGCCGCGTCGCCACTTG 1904
Qy 1923 GCCTGGATGGTGAATGGGTGATGGCCTGCACAAATTCAGCAGTGGCCTAGCAATGGT 1982
Db 1905 GCCTGGATGGTGAATGGGTGATGGCCTGCACAAATTCAGCAGTGGCCTAGCAATGGT 1964
Qy 1983 GCTGCTTTTACTGAAGCTTTATCAAGTGGTTTAAAGTACTTCTGCTGCTGTTCTGTCAT 2042
Db 1965 GCTGCTTTTACTGAAGCTTTATCAAGTGGTTTAAAGTACTTCTGCTGCTGTTCTGTCAT 2024
Qy 2043 GAGTTGCTCATGAATAGTGAAGTCTGCTGCTTCTACTAAAGCTGGCATGACCGTTAAG 2102
Db 2025 GAGTTGCTCATGAATAGTGAAGTCTGCTGCTTCTACTAAAGCTGGCATGACCGTTAAG 2084
Qy 2103 CAGGCTGCTCTTTATAATGCAATGTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGA 2162
Db 2085 CAGGCTGCTCTTTATAATGCAATGTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGA 2144
Qy 2163 ATTTTTCATGTCATATGCTGCAAAATCTTCTATGTTGGATATTTGACATTTACTGCTGGC 2222
Db 2145 ATTTTTCATGTCATATGCTGCAAAATCTTCTATGTTGGATATTTGACATTTACTGCTGGC 2204
Qy 2223 TTATTCATGATGTTGCTGCTGCTGATGATGCTACCTGAAATGCTGCAAAATGATGCTGCT 2282
Db 2205 TTATTCATGATGTTGCTGCTGCTGATGATGCTACCTGAAATGCTGCAAAATGATGCTGCT 2264
Qy 2283 GACCATGATGATGAGCCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTGGGT 2342
Db 2265 GACCATGATGATGAGCCGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTGGGT 2324

Qy 2343 TTTGGAATTTAGTTTACTTATTTCCATATTTTGAACATAAAAATCGTGTTCGTATAAAATTC 2402
Db 2325 TTTGGAATTTAGTTTACTTATTTCCATATTTTGAACATAAAAATCGTGTTCGTATAAAATTC 2384
Qy 2403 TAGTTAAGTTTAAATGCTAGTAGCTTAAAGAGTTGTATAGTTTTCAGTAGTGTATAG 2462
Db 2385 TAGTTAAGTTTAAATGCTAGTAGCTTAAAGAGTTGTATAGTTTTCAGTAGTGTATAG 2444
Qy 2463 GGAGATGAGTTTGTATGCTGCTACTATGACGCTTTAAAGTTAGTGGGTTTGTGATTTT 2522
Db 2445 GGAGATGAGTTTGTATGCTGCTACTATGACGCTTTAAAGTTAGTGGGTTTGTGATTTT 2504
Qy 2523 GTATTGAATTTGCTGCTGCTGTACAAAGTCAGTTAAAGGTACGTTTAAATATTTAAGCTA 2582
Db 2505 GTATTGAATTTGCTGCTGCTGTACAAAGTCAGTTAAAGGTACGTTTAAATATTTAAGCTA 2564
Qy 2583 TTTCTATCTTGGAGATTAATCTGTATGTGCAATTCACCGGTATTTACCAGTTTATATGTA 2642
Db 2565 TTTCTATCTTGGAGATTAATCTGTATGTGCAATTCACCGGTATTTACCAGTTTATATGTA 2624
Qy 2643 AACAGAGATTGGCATGACATGTTCTGTATGTTTCAGGGGAAAAATGCTTTTAATGCTTT 2702
Db 2625 AACAGAGATTGGCATGACATGTTCTGTATGTTTCAGGGGAAAAATGCTTTTAATGCTTT 2684
Qy 2703 TTCAGAACTAACACAGTTATTTCTTATCTGATTTTAGTCTCTGAAAGAACTGCTGTG 2762
Db 2685 TTCAGAACTAACACAGTTATTTCTTATCTGATTTTAGTCTCTGAAAGAACTGCTGTG 2744
Qy 2763 TTTTAGGAATAAGATTGTCATGAAGCTTAAATACCAAGAAAGCTTATCTGNAATTTAAG 2822
Db 2745 TTTTAGGAATAAGATTGTCATGAAGCTTAAATACCAAGAAAGCTTATCTGNAATTTAAG 2804
Qy 2823 CAAAGAAATAAGAGGAGAAAGAGAAATCTGAGAAATTTGGGAGGCGATAGATCTTATAA 2882
Db 2805 CAAAGAAATAAGAGGAGAAAGAGAAATCTGAGAAATTTGGGAGGCGATAGATCTTATAA 2864
Qy 2883 AAATCACAAAATTTGTTGTAATTTAGAGGGAGAAATTTAGAAATTAAGTATATAAAGGCA 2942
Db 2865 AAATCACAAAATTTGTTGTAATTTAGAGGGAGAAATTTAGAAATTAAGTATATAAAGGCA 2924
Qy 2943 GAATTAGTAGTAGTACATTCATTAACATTTTGTGAGGATTTTCCCGTAAAGAGCT 3002
Db 2925 GAATTAGTAGTAGTACATTCATTAACATTTTGTGAGGATTTTCCCGTAAAGAGCT 2984
Qy 3003 AGTGAGCAC-TCTCATATACTAAT--AGTGACATTTAACTTTGTATAATACAGAAATC 3059
Db 2985 AGTGAGCAC-TCTCATATACTAAT--AGTGACATTTAACTTTGTATAATACAGAAATC 3044
Qy 3060 TAAATATATTTAATGAATTCAGCAATATA-CACCTTGACCAAGAAATTTGGAATTTCAAAA 3118
Db 3045 TAAATATATTTAATGAATTCAGCAATATA-CACCTTGACCAAGAAATTTGGAATTTCAAAA 3104
Qy 3119 TGTTCGCGGGTTATATACAGATGAGTACAGTAGTAG-TTTATGATACACAGACTG 3177
Db 3105 TGTTCGCGGG-TATATACAGATGAGTACAGTAGTAGTTTATGATACACAGACTG 3163
Qy 3178 GGTATTGCGCAAGTTATATACCAAAAGCTGTATGACTGGATGTTCTGTTTACCTGCT 3237
Db 3164 GGTATTGCGCAAGTTATATACCAAAAGCTGTATGACTGGATGTTCTGTTTACCTGCT 3223
Qy 3238 TTAACAAATTTATCAGAGTAGTAAACTTTGATATATATGAGGATATTAACACTACACTAA 3297
Db 3224 TTAACAAATTTATCAGAGTAGTAAACTTTGATATATATGAGGATATTAACACTACACTAA 3283
Qy 3298 GTATCATTTGATTCGATTTCAGAAAGTACTTTGATATCTCTCAGTGTCTCAGTGTATCAT 3357
Db 3284 GTATCATTTGATTCGATTTCAGAAAGTACTTTGATATCTCTCAGTGTCTCAGTGTATCAT 3343
Qy 3358 TGTGAGCAATTTGCT-TTTATATACGGTACTGTAGCCATACTAGGCCCTGCTGTGGCATTC 3416
Db 3344 TGTGAGCAATTTGCTTTTATATACGGTACTGTAGCCATACTAGGCCCTGCTGTGGCATTC 3403

QY 3417 TCTAGATGTTTCTTTTACACAAATAAATTCCTTATATCAGC 3458
DB 3404 TCTAGATGTTTCTTTTACACAAATAAATTCCTTATATCAGC 3445

RESULT 5
AADI3480
ID AAD13480 standard; cDNA; 2776 BP.
XX AAD13480;
AC
XX
XX 06-NOV-2001 (first entry)
XX Human LIV-1-164647 cDNA.
XX
XX Human; LIV-1-164647; cytostatic; estrogen-inducible gene; tumour;
KW cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;
KW salivary gland; carcinoma; drug screening; therapy; ss.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH CDS 73..2340
FT /*tag= a
FT /product= "Human LIV-1-164647 protein"
XX
XX WO200155178-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02622.
XX
XX 25-JAN-2000; 2000US-0177951.
PR 10-APR-2000; 2000US-0195761.
XX
XX (GETH) GENENTECH INC.
XX
XX Goddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;
PI WPI: 2001-502628/55.
DR P-PSDB; AAE07206.
XX
XX New LIV-1 DNA and amino acids, useful for treating cancer or inhibiting
PT the proliferation of tumor cells in mammals, e.g. breast, lung,
PT prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma
PT
XX
XX Claim 6; Fig 2A; 150pp; English.
XX
XX The present sequence is human estrogen-inducible LIV-1-164647
CC cDNA. LIV-1 is overexpressed in tumour tissues such as prostate,
CC colon, lung and breast. The LIV-1 DNA and polypeptide, and the monoclonal
CC antibody are useful for treating cancer and inhibiting the proliferation
CC of tumour cells in mammals, e.g. breast, lung, prostate, colon, ovary,
CC uterus, kidney, gastric or salivary carcinoma, or other tumour cell
CC types expressing the LIV-1-164647 protein. In particular, the mammal is
CC a human. The LIV-1 DNA and polypeptide may also be used in screening
CC assays for drug candidates.
XX
XX Sequence 2776 BP; 823 A; 583 C; 592 G; 778 T; 0 other;
SQ

Query Match 79.8%; Score 2761.2; DB 22; Length 2776;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2766; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 68 GCCCGTGTGAACCAACCTGCGCGGTGGCCGGCGGTGGGACACAGGCGCGGGAGA 127
DB 1
3 GGCCGTGTGAACCAACCTGCGCGGTGGCCGGCGGTGGGACACAGGCGCGGGAGA 62

QY 128 CGAAGGCGCAATGCCGAGGAAGTTATCTGTAATCTTGATCCCTGACCTTTGCCCTCTCTGT 187
DB 63 CGAAGGCGCAATGCCGAGGAAGTTATCTGTAATCTTGATCCCTGACCTTTGCCCTCTCTGT 122

QY 188 CACAAATCCCCTTCTATGAACATAAAGCAGCTGCTTTCCCCCAGACACTGAGAAAATTAG 247
DB 123 CACAAATCCCCTTCTATGAACATAAAGCAGCTGCTTTCCCCCAGACACTGAGAAAATTAG 182

QY 248 TCCGAATTGGGAATCTGGCATTAAATGTTGACTTGGCAATTTCCACAGGCAATATCATCT 307
DB 183 TCCGAATTGGGAATCTGGCATTAAATGTTGACTTGGCAATTTCCACAGGCAATATCATCT 242

QY 308 ACAACAGCTTTTCTACCGCTATGGAGAAAATAAATTTCTTTGTGCTTGAAGGTTTCAGAAA 367
DB 243 ACAACAGCTTTTCTACCGCTATGGAGAAAATAAATTTCTTTGTGCTTGAAGGTTTCAGAAA 302

QY 368 ATTACTTCAAAATATAGGCATTAGATAAGATTAAAGAAATCCATATACACCATGACCAGA 427
DB 303 ATTACTTCAAAATATAGGCATTAGATAAGATTAAAGAAATCCATATACACCATGACCAGA 362

QY 428 CCATCCTCAGACACGAGCATCTCAGACCATGAGCGTCACTCAGACCATGAGCATCA 487
DB 363 CCATCCTCAGACACGAGCATCTCAGACCATGAGCGTCACTCAGACCATGAGCATCA 422

QY 488 CTCAGACACGAGCATCTCTGACCATGATCATCTCTCACCATATCATGCTGCTTC 547
DB 423 CTCAGACACGAGCATCTCTGACCATGATCATCTCTCACCATATCATGCTGCTTC 482

QY 548 TGGTAAAAATTAAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTCCAGTAAAGA 607
DB 483 TGGTAAAAATTAAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTCCAGTAAAGA 542

QY 608 TCCTAGAAACGCCAGGGAAGAGCTCACCGACCAAGCATGCCAGTGGTAGAAGGAA 667
DB 543 TCCTAGAAACGCCAGGGAAGAGCTCACCGACCAAGCATGCCAGTGGTAGAAGGAA 602

QY 668 TGTCAAGGACAGTGTGTAGTGTAGTGAAGTGACCTCAACTGTGTACAACTGTCTCTGA 727
DB 603 TGTCAAGGACAGTGTGTAGTGTAGTGAAGTGACCTCAACTGTGTACAACTGTCTCTGA 662

QY 728 AGGAATCACTTTCTAGAGACAATAGAGACTCCAAAGACTCGAAACCTTTCCCCAAAGA 787
DB 663 AGGAATCACTTTCTAGAGACAATAGAGACTCCAAAGACTCGAAACCTTTCCCCAAAGA 722

QY 788 TGTAAAGAGCTCCACTCCACCAGTGTGCATCAAAAGCCGGGTGAGCCGGCTGGTGG 847
DB 723 TGTAAAGAGCTCCACTCCACCAGTGTGCATCAAAAGCCGGGTGAGCCGGCTGGTGG 782

QY 848 TAGGAAAACAATGAAATCTGTGAGTGGAGCCCGGAAAGGCTTTATGTATTCAGAAACAC 907
DB 783 TAGGAAAACAATGAAATCTGTGAGTGGAGCCCGGAAAGGCTTTATGTATTCAGAAACAC 842

QY 908 AAATGAAAATCCTCAGGAGTGTTCATGCAATCAAAAGCTACTGCATCTCATGCGATGGG 967
DB 843 AAATGAAAATCCTCAGGAGTGTTCATGCAATCAAAAGCTACTGCATCTCATGCGATGGG 902

QY 968 CATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTGTGCCAGCATCATCAACCA 1027
DB 903 CATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTGTGCCAGCATCATCAACCA 962

QY 1028 AATTGATGCTAGATCTCTCTGATTATCAAGTGAAAGAGGCTGAAATCCCTCCAAA 1087
DB 963 AATTGATGCTAGATCTCTCTGATTATCAAGTGAAAGAGGCTGAAATCCCTCCAAA 1022

QY 1088 GACCTATTTCATTACAAATAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1147
DB 1023 GACCTATTTCATTACAAATAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1082

QY 1148 CTTGTCTCTGTGGGGGTTATCTTAGTGCCTCTCATGAAATCGGGTGGTGGTGGTGGTGG 1207
DB 1083 CTTGTCTCTGTGGGGGTTATCTTAGTGCCTCTCATGAAATCGGGTGGTGGTGGTGGTGG 1142

QY 1208 CTTGAGTTTCTTGTGGGCACTGGCGGTTGGGACTTTGAGTGGTGGTGGTGGTGGTGGTGG 1267
DB 1143 CTTGAGTTTCTTGTGGGCACTGGCGGTTGGGACTTTGAGTGGTGGTGGTGGTGGTGGTGG 1202

QY 1268 TCTTCCCAATTCCTCATGCAAGTCCACCACCATAGTATAGCCATGAAGAACCAGCAATGGA 1377

Db 1203 TCTTCCACATCTCATGCAAGTCCACCACATAGTCATAGCATGCAAGAACCCACATGGA 1262
 Qy 1328 AATGAAAGAGGACCACTTTTCAGTCATCTGCTCTTCAAAACATAGAAGAAAGTGCCCTA 1387
 Db 1263 AATGAAAGAGGACCACTTTTCAGTCATCTGCTCTTCAAAACATAGAAGAAAGTGCCCTA 1322
 Qy 1388 TTTTGTATTCACGCTGGAAGGGTCTAACAGCTCTAGGAGGCTGTATTTTCATGTTCTTGT 1447
 Db 1323 TTTTGTATTCACGCTGGAAGGGTCTAACAGCTCTAGGAGGCTGTATTTTCATGTTCTTGT 1382
 Qy 1448 TGAACATGCTCCATCATCAATCATCAACAAATTTAAAGATAGAAGAAAGAAATCAGAAGAA 1507
 Db 1383 TGAACATGCTCCATCATCAATCATCAACAAATTTAAAGATAGAAGAAAGAAATCAGAAGAA 1442
 Qy 1508 ACCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1567
 Db 1443 ACCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1502
 Qy 1568 TTAACAAATGAGGAGAAAGTACATACAGATGATGATGATGATGATGATGATGATGATGATGAT 1627
 Db 1503 TTAACAAATGAGGAGAAAGTACATACAGATGATGATGATGATGATGATGATGATGATGATGAT 1562
 Qy 1628 CTCAAGAGGCTCCACCTTTGATTTCTCAGAGCTGCTGGAAGAAAGAGGT 1687
 Db 1563 CTCACAGAGGCTCCACCTTTGATTTCTCAGAGCTGCTGGAAGAAAGAGGT 1622
 Qy 1688 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1747
 Db 1623 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1682
 Qy 1748 GAATTAATGCCATTTCCACATTTCCACATTTCCACATTTCCACATTTCCACATTTCCACAT 1807
 Db 1683 GAATTAATGCCATTTCCACATTTCCACATTTCCACATTTCCACATTTCCACATTTCCACAT 1742
 Qy 1808 CCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867
 Db 1743 CCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1802
 Qy 1868 TCACAGCAGGCTACTCTCGGAGGAGCTGAAAGATGCGGGGCTGCCACTTTGGCCTG 1927
 Db 1803 TCACAGCAGGCTACTCTCGGAGGAGCTGAAAGATGCGGGGCTGCCACTTTGGCCTG 1862
 Qy 1928 GAT 1987
 Db 1863 GAT 1922
 Qy 1988 TTTTACTGAAGCTTATCAAGTGGTTAAGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2047
 Db 1923 TTTTACTGAAGCTTATCAAGTGGTTAAGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1982
 Qy 2048 GCCTCATGAATAGGTGACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2107
 Db 1983 GCCTCATGAATAGGTGACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2042
 Qy 2108 TGTCTCTTATAATGATGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2167
 Db 2043 TGTCTCTTATAATGATGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2102
 Qy 2168 CATTTGCTCATTTGCTGAAATGTTTCTATGCTGATTTTTCATTTTTCATTTTTCATTTTTCATTTT 2227
 Db 2103 CATTTGCTCATTTGCTGAAATGTTTCTATGCTGATTTTTCATTTTTCATTTTTCATTTTTCATTTT 2162
 Qy 2228 CATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2287
 Db 2163 CATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2222
 Qy 2288 TGGATGTAGCGCTGGGGGTATTTCTTTTACAGATGCTGGGATGCTTTTGGGTTTGG 2347
 Db 2223 TGGATGTAGCGCTGGGGGTATTTCTTTTACAGATGCTGGGATGCTTTTGGGTTTGG 2282
 Qy 2348 AATTATGTTACTTATTTCCATATTTTGAACATATAAATCGTCTTTCGTATATAAATTTCTAGTT 2407

Db 2283 AATTATGTTACTTATTTCCATATTTTGAACATATAAATCGTCTTTCGTATAAATTTCTAGTT 2342
 Qy 2408 AAGGTTTAAATGCTAGAGTAGCTTTAAAGAGTTGTGCATAGTTTTCAGTAGGTCATAGGAGA 2467
 Db 2343 AAGGTTTAAATGCTAGAGTAGCTTTAAAGAGTTGTGCATAGTTTTCAGTAGGTCATAGGAGA 2402
 Qy 2468 TGAGTTTGTATGCTGCTACTATGCGAGCGTTTAAAGCTTAGTGGGTTTTCGTGATTTTCTATT 2527
 Db 2403 TGAGTTTGTATGCTGCTACTATGCGAGCGTTTAAAGCTTAGTGGGTTTTCGTGATTTTCTATT 2462
 Qy 2528 GAATATTGCTGCTGCTTACAAAGTCAGTTAAAGGTACGTTTTTAATATTTTAAGTTATTCTA 2587
 Db 2463 GAATATTGCTGCTGCTTACAAAGTCAGTTAAAGGTACGTTTTTAATATTTTAAGTTATTCTA 2522
 Qy 2588 TCTTGAGATATAATGCTGCTATGTCGAATTCACCGGTATTACCAAGTTTATTATGTAACAA 2647
 Db 2523 TCTTGAGATATAATGCTGCTATGTCGAATTCACCGGTATTACCAAGTTTATTATGTAACAA 2582
 Qy 2648 GAGATTTGGCATGACATGTTCTGTATGTTTTCAGGGAAATGCTTTAATGCTTTTCA 2707
 Db 2583 GAGATTTGGCATGACATGTTCTGTATGTTTTCAGGGAAATGCTTTAATGCTTTTCA 2642
 Qy 2708 GAACTAACACAGTTTTCCTATACCTGATTTTAGGCTCTCTGAAGAACTGCTGCTGTTTAG 2767
 Db 2643 GAACTAACACAGTTTTCCTATACCTGATTTTAGGCTCTCTGAAGAACTGCTGCTGTTTAG 2702
 Qy 2768 GAACTAACAGTGTGCATGAAGCTTAAATACCAAGAAAGCTTATACGAAATTTAAGCAAG 2827
 Db 2703 GAACTAACAGTGTGCATGAAGCTTAAATACCAAGAAAGCTTATACGAAATTTAAGCAAG 2762
 Qy 2828 AATTAAGAGGAGAA 2841
 Db 2763 AAAAAAAAAAAAAA 2776

RESULT 6
 AAT99071 standard; cDNA; 2404 BP.
 ID AAT99071 standard; cDNA; 2404 BP.
 XX AAT99071:
 XX 24-MAR-1998 (first entry)
 XX Partial sequence of the pLIV1 gene.
 DE
 XX Oestrogen regulated gene; pLIV1 gene partial sequence; breast cancer;
 KW metastatic risk assessment; hormone therapy responsiveness;
 KW tumour oestrogen receptor expression; lymph node involvement; ds.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FH 1..1296
 FT /*tag= a
 XX US5693465-A.
 XX 02-DEC-1997.
 XX 22-SEP-1994; 94US-0311023.
 XX 22-SEP-1994; 94US-0311023.
 XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
 XX Gee JMW, Green CD, Manning DL, Nicholson RI;
 PI WPI; 1998-031740/03.
 DR P-PSDB; AAW34528.
 XX Assessment of metastatic risk or oestrogen responsiveness in breast
 PT cancer - by detecting expression of pLIV1 protein
 XX

PS Disclosure: Column 11-14; 11pp; English.

This sequence represents a partial sequence of the pLIV1 gene. This sequence extends further in the 3' direction than the pLIV1 partial sequence shown in AAT99070. The protein encoded by this sequence is an oestrogen regulated gene associated with breast cancer. The protein encoded by this sequence can be used in the method of the invention. The method is for determining the risk of metastasis of a female breast tumour and predicting the responsiveness of a female breast tumour to hormone treatment comprise determining if a tissue sample from the tumour expresses a polypeptide comprising at least 14 continuous amino acids of the protein encoded by this sequence. The method is useful for determining the risk of metastasis of a female breast tumour and to predict the responsive of the female breast tumour to hormone treatment. pLIV1 expression is highly predictive of tumour oestrogen receptor expression and lymph node involvement.

Sequence 2404 BP; 726 A; 411 C; 492 G; 775 T; 0 other;

Query Match	65.0%	Score 2249;	DB 19;	Length 2404;
Best Local Similarity	97.4%	Pred. No. 0;		
Matches 2341; Conservative	0;	Mismatches	0;	Indels 62; Gaps 3;

Qy	1119	GGTTTTATAGCCATTTCCATCATCAGTTTCTCTGCTCTGCTGGGGGTATCTCTTAGTGCCT	1178
Db	1	GGTTTTATAGCCATTTCCATCATCAGTTTCTCTGCTCTGCTGGGGGTATCTCTTAGTGCCT	60
Qy	1179	CTCATCAATCGGGTGTTTTTCAAAATTTCTCCTGAGTTTCCCTTGCGCACATGGCCGTTGGG	1238
Db	61	CTCATGANTCGGGTGTTTTTCAAAATTTCTCCTGAGTTTCTCTGTTGGCACATGGCCGTTGGG	120
Qy	1239	ACTTTGAGTGGTGATGCTTTTTTACACCTCTCTTCACATTTCTCATGCAAGTCACCAACAT	1298
Db	121	ACTTTGAGTGGTGATGCTTTTTTACACCTCTCTTCACATTTCTCATGCAAGTCACCAACAT	180
Qy	1299	AGTCATAGCCATGAGAACCAGCAATGGAATGAAAGAGACACACTTTTCAGTCATCTG	1358
Db	181	AGTCATAGCCATGAGAACCAGCAATGGAATGAAAGAGACCACTTTTCAGTCATCTG	240
Qy	1359	TCCTTCTCAAAACATAGAAGAAAGTGCCTATTTTGATTTCCACGTGGAAGGGTCTAACAGCT	1418
Db	241	TCCTTCTCAAAACATAGAAGAAAGTGCCTATTTTGATTTCCACGTGGAAGGGTCTAACAGCT	300
Qy	1419	CTAGGAGGCGCTGATTTTCATGTTTCTGTTGAACATGTCTTCACATGTGATCAACAATTT	1478
Db	301	CTAGGAGGCGCTGATTTTCATGTTTCTGTTGAACATGTCTTCACATGTGATCAACAATTT	360
Qy	1479	AAAGATAGAAGAAAAGAAATCAGAAGAAACCTGAAAATGATGATGTGGAGATTAAAG	1538
Db	361	AAAGATAGAAGAAAAGAAATCAGAAGAAACCTGAAAATGATGATGTGGAGATTAAAG	420
Qy	1539	AAGCATTTGTCCAAGTATGAAATCTCAACTTTTCAACAAATGAGGAGAAAGTAGATACAGAT	1598
Db	421	AAGCATTTGTCCAAGTATGAAATCTCAACTTTTCAACAAATGAGGAGAAAGTAGATACAGAT	480
Qy	1599	GATCGAATGAGGCTATTTACGACGACACTCACAAAGAGCCCTCCCACTTTGATCTTCAG	1658
Db	481	GATCGAATGAGGCTATTTACGACGACACTCACAAAGAGCCCTCCCACTTTGATCTTCAG	540
Qy	1659	CAGCCTGCGAGTCTTGGAAAGAAGAGGTTCATGATAGCTCATGTCTCATCCACAGGAAATC	1718
Db	541	CAGCCTGCGAGTCTTGGAAAGAAGAGGTTCATGATAGCTCATGTCTCATCCACAGGAAATC	600
Qy	1719	TACAATGAATATGTATCCAGAGGGTGCAAGAAATAATGCCATTCACATTTCCACGATACA	1778
Db	601	TACAATGAATATGTATCCAGAGGGTGCAAGAAATAATGCCATTCACATTTCCACGATACA	660
Qy	1779	CTCGGCCAGTCAGACGATCTCAATCACCACCATATGACTACCATCATATTTCTCCATCAT	1838
Db	661	CTCGGCCAGTCAGACGATCTCAATCACCACCATATGACTACCATCATATTTCTCCATCAT	720
Qy	1839	CACCACCCAAAACCAACCATCCTTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGGCTG	1898

Db	601	TACAA	TGAAT	ATGT	ACC	CAGAG	GGTG	CAAG	ATAA	TGCA	TTAC	ACAT	TTCC	ACG	ATA	CA	660
QY	1779	CTCG	CCAG	TCT	CA	TTC	ATC	AC	ATCAT	CA	TG	AGC	ATCAT	CA	TAT	TC	1838
Db	661	CTCG	CCAG	TCT	CA	TTC	ATC	AC	ATCAT	CA	TG	AGC	ATCAT	CA	TAT	TC	720
QY	1839	CAC	CA	CC	AA	AC	CA	CC	AT	CT	CA	CA	GT	CA	CA	CC	1898
Db	721	CAC	CA	CC	AA	AC	CA	CC	AT	CT	CA	CA	GT	CA	CA	CC	780
QY	1899	AA	GA	T	CG	CG	CT	CG	CA	CT	TT	GG	CG	TG	AT	GG	1958
Db	781	AA	GA	T	CG	CG	CT	CG	CA	CT	TT	GG	CG	TG	AT	GG	840
QY	1959	TT	CA	GG	AT	GC	CT	TA	GC	AA	TT	GG	T	CT	GA	AG	2018
Db	841	TT	CA	GG	AT	GC	CT	TA	GC	AA	TT	GG	T	CT	GA	AG	900
QY	2019	AC	TT	CT	GT	CT	GT	CT	CA	TG	AG	TT	GC	CT	CA	TA	2078
Db	901	AC	TT	CT	GT	CT	GT	CT	CA	TG	AG	TT	GC	CT	CA	TA	960
QY	2079	CT	AA	AG	CG	TG	CA	TG	AG	CG	TG	CT	TT	AT	AT	AT	2138
Db	961	CT	AA	AG	CG	TG	CA	TG	AG	CG	TG	CT	TT	AT	AT	AT	1020
QY	2139	CG	TA	CT	TT	CG	TA	CT	TT	CG	TA	CT	TT	CG	TA	CT	2198
Db	1021	CG	TA	CT	TT	CG	TA	CT	TT	CG	TA	CT	TT	CG	TA	CT	1080
QY	2199	TG	GA	TA	TT	GC	TA	TT	GC	TA	TT	GC	TA	TT	GC	TA	2258
Db	1081	TG	GA	TA	TT	GC	TA	TT	GC	TA	TT	GC	TA	TT	GC	TA	1140
QY	2259	GA	AT	GC	TG	CA	CA	AT	GA	TG	CA	CA	AT	GA	TG	CA	2318
Db	1141	GA	AT	GC	TG	CA	CA	AT	GA	TG	CA	CA	AT	GA	TG	CA	1200
QY	2319	CA	GA	AT	GCT	GG	AT	GCT	TT	TG	GG	AT	TG	GA	AT	TG	2378
Db	1201	CA	GA	AT	GCT	GG	AT	GCT	TT	TG	GG	AT	TG	GA	AT	TG	1259
QY	2379	AA	AT	CC	GT	TT	CG	TA	AA	TT	TC	TA	GT	TT	AA	AG	2438
Db	1260	AA	AT	CC	GT	TT	CG	TA	AA	TT	TC	TA	GT	TT	AA	AG	1318
QY	2439	TG	CA	TA	GT	TT	CA	GT	AG	CA	TA	GT	TT	CA	GT	AG	2498
Db	1319	TG	CA	TA	GT	TT	CA	GT	AG	CA	TA	GT	TT	CA	GT	AG	1378
QY	2499	AA	GT	TAG	TGG	GG	TT	TG	TA	TG	AA	TT	TG	TA	TG	AA	2558
Db	1379	AA	GT	TAG	TGG	GG	TT	TG	TA	TG	AA	TT	TG	TA	TG	AA	1438
QY	2559	AG	GT	TAC	CT	TT	TA	AT	TT	AA	GT	TT	CT	AT	CT	TT	2618
Db	1439	AG	GT	TAC	CT	TT	TA	AT	TT	AA	GT	TT	CT	AT	CT	TT	1498
QY	2619	CC	GG	TAT	TAC	CA	GT	TT	AT	TG	TA	AA	CA	GT	TT	GC	2678
Db	1499	CC	GG	TAT	TAC	CA	GT	TT	AT	TG	TA	AA	CA	GT	TT	GC	1558
QY	2679	AG	GG	AA	AT	GT	CT	TT	TA	TG	CT	TT	TA	AG	AA	CT	2738
Db	1559	AG	GG	AA	AT	GT	CT	TT	TA	TG	CT	TT	TA	AG	AA	CT	1618
QY	2739	TAG	GT	CT	CT	GA	AA	CT	GCT	GT	GT	TT	TG	GA	AT	TG	2798
Db	1619	TAG	GT	CT	CT	GA	AA	CT									

QY	2859	TTGGGAGGCATAGATCTTTATAAAATCACAAAATTTGTTGTAATATAGAGGGGAGAAA	2918
Db	1739	TTGGGAGGCATAGATCTTTATAAAATCACAAAATTTGTTGTAATATAGAGGGGAGAAA	1799
QY	2919	TTTACAAATTAAGTATAAAAAGGCAGAAATAGTATACAGTACATTCATTAACATTTTTGT	2978
Db	1799	TTTACAAATTAAGTATAAAAAGGCAGAAATAGTATACAGTACATTCATTAACATTTTTGT	1858
QY	2979	CAGGAATATTTCCCGTAAAAACGTAAGTACGACCTCTCATATACTAATAGTGACATTTTA	3038
Db	1859	CAGGAATATTTCCCGTAAAAACGTAAGTACGACCTCTCATATACTAATAGTGACATTTTA	1918
QY	3039	ACTTTGTATAATACAGAAATCTAAATATATTTAATGAATCAAGCAATATACACTTGACC	3098
Db	1919	ACTTTGTATAATACAGAAATCTAAATATATTTAATGAATCAAGCAATATACACTTGACC	1978
QY	3099	AAGAAATTTGGAATTTCAAAATGTTCTGTCGGGTTATATACCAGATGAGTACAGTCAGTAG	3158
Db	1979	AAGAAATTTGGAATTTCAAAATGTTCTGTCGGGTTATATACCAGATGAGTACAGTCAGTAG	2038
QY	3159	TTTATGTATCACCAGACTGGGTATTGCGCAAGTTATATATCACAAAAGCTGTATGACTG	3218
Db	2039	TTTATGTATCACCAGACTGGGTATTGCGCAAGTTATATATCACAAAAGCTGTATGACTG	2098
QY	3219	GATGCTTCCTGGTTACTCGTTTACAAAATATATCAGAGTACT-----	3258
Db	2099	GATGCTTCCTGGTTACTCGTTTACAAAATATATCAGAGTACT-----	2158
QY	3259	-----AAAACCTTTCATATATATCAG	3278
Db	2159	GATATTAACACTACACTAAGTATCATTTGATTCGATTTCAGAAAACCTTTGATATATATGAG	2218
QY	3279	GATATTAACACTACACTAAGTATCATTTGATTCGATTTCAGAAAAGTACTTTGATATCTCTC	3338
Db	2219	GATATTAACACTACACTAAGTATCATTTGATTCGATTTCAGAAAAGTACTTTGATATCTCTC	2278
QY	3339	AGTGCTTCAGTGCTATCATTTGTGAGCAATGTCTTTATATACGGTACTGTAGCCATACCTA	3398
Db	2279	AGTGCTTCAGTGCTATCATTTGTGAGCAATGTCTTTATATACGGTACTGTAGCCATACCTA	2338
QY	3399	GGCTGTCTGGGCATCTACATGTTTCTTTTACACAATAAATTCCTTTATATCAGC	3458
Db	2339	GGCTGTCTGGGCATCTACATGTTTCTTTTACACAATAAATTCCTTTATATCAGC	2398
QY	3459	TTG 3461	
Db	2399	TTG 2401	

RESULT 8
AAZ33622
ID AAZ33622 standard; cDNA; 1597 BP.
XX
AC AAZ33622;
XX
XX
DT 08-DEC-1999 (first entry)
XX
XX
DE Human breast tumour-associated EST 12.
XX
XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;
KW treatment; tumour; cytostatic; medication; ss.
KW

RESULT. T. 8

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;
XX
DR WPI; 1999-528981/45.

Human nucleic acid sequences and protein products from tumor breast tissue, useful for breast cancer therapy -

PS Claim 3; 93; 188pp; German.

This invention describes novel human nucleic acid sequences from tumor breast tissue which have cytostatic activity. The nucleic acid sequences can be used to produce and isolate full-length gene sequences. They can be used to express proteins, which can be used as tools to find an activity against breast cancer. The sequences can be used in sense or antisense form. They are especially useful for medications for gene therapy to treat breast cancer. AZ33611-248617 represents expressed sequence tags described in the method of the invention.

Sequence 1597 BP; 476 A; 229 C; 337 G; 555 T; 0 other;

Query Match 43.0%; Score 1486.8; DB 20; Length 1597;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1559; Conservative 0; Mismatches 7; Indels 8; Gaps 6;

Qy 1895 GCTGAAAGATGCCGGCGTCGCCACTTTGGCCTGGATGGTGATATATGGTGATGGCCCTGCA 1954

Db 19 GCTGAAAGATGCCGGCGTCGCCACTCTGGCCTGGATGGTGATATATGGTGATGGCCCTGCA 78

Qy	1955	CAATTT	CAGCGAT	GGCCT	AGCAAT	TGGTGC	TTTAC	TGAAG	GCTTAT	CAAGT	GCTTT	2014
Db	79	CAATTT	CAGCGAT	GGCCT	AGCAAT	TGGTGC <td>TTTAC <td>TGAAG <td>GCTTAT <td>CAAGT <td>GCTTT</td> <td>138</td> </td></td></td></td>	TTTAC <td>TGAAG <td>GCTTAT <td>CAAGT <td>GCTTT</td> <td>138</td> </td></td></td>	TGAAG <td>GCTTAT <td>CAAGT <td>GCTTT</td> <td>138</td> </td></td>	GCTTAT <td>CAAGT <td>GCTTT</td> <td>138</td> </td>	CAAGT <td>GCTTT</td> <td>138</td>	GCTTT	138

Oy	2015	AAGTACTTCTGTGCTGTCTCTCATGAGTTGCTCATGAATTAGGTGACTTTGCTGT	2074
Db	139	AAGTACTTCTGTGCTGTCTCTCATGAGTTGCTCATGAATTAGGTGACTTTGCTGT	198

Qy	2075	TCTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGTCCTTTATATGTCATTGTCAGGCAT	2134
Db	199	TCTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGTCCTTTATATGTCATTGTCAGGCAT	258

QY 2135 GCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTGGTCATTATGCTGAAATGTTTC 2194
|||||
Db 259 GCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTGGTCATTATGCTGAAATGTTTC 318

Qy	2195	TATGTCGATATTGCCACTTACTGCTGGCTATTTCATGTATGTTGCCTGGTTGATATGGT	2254
Db	319	TATGTCGATATTGCCACTTACTGCTGGCTATTTCATGTATGTTGCCTGGTTGATATGGT	378

QY	2255	ACCTGAAATGCTGCACAATGATGCTAGTGACCATTGGATGTAGCCGCTGGGGTATTTCCT	2314
Db	379	ACCTGAAATGCTGCACAATGATGCTAGTGACCATTGGATGTAGCCGCTGGGGTATTTCCT	438

QY 2315 TTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTGA 2374
|||||
Db 439 TTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTGA 498

Qy 2375 ACATAAAATCGTGTTCGCTATAAATTTCTAGTTAAGGTTTAAATGCTAGAGTAGCTTAAA 2434
|||||
Db 499 ACATAAAATCGGTTCGCTATAAATTTCTAGTTAAGGTTTAAATGCTAGAGTAGCTTAAA 558

QY 2435 AAGTTGCTCATAGTTTCAGTAGGTCATAGGAGATGAGTTTGTATGCTGTACTATGCAGCG 2494
|||||
Db 559 AAGTTGCTCATAGTTTCAGTAGGTCATAGGAGATGAGTTTGTATGCTGTACTATGCAGCG 618

QY	2495	TTTAAAGTTAGTGGGTTTTGTGATTTTGTATTGAATAATTGCTGCTGTATACAAAGTCAG	2554
Db	619	TTTAAAGTTACTGGGTTTTGTGATTTTGTATTGAATAATTGCTGCTGTATACAAAGTCAG	678

QY	2555	TTAAAGGTACGTTT	TAATATTT	AAGTTATCT	ATCTTGGAGAT	AAATCTG	TATGTGCAA	2614
Db	679	TTAAAGGTACGTTT <th>TAATATTT</th> <th>AAGTTATCT</th> <th>ATCTTGGAGAT</th> <th>AAATCTG</th> <th>TATGTGCAA</th> <th>738</th>	TAATATTT	AAGTTATCT	ATCTTGGAGAT	AAATCTG	TATGTGCAA	738

QY 2615 TTCACCGGTATTACCAAGTTTATTATGTAACAAGAGATTTGGCATGACATGTTCTGTATG 2674
|||||
Db 739 TTCACCGGTATTACCAAGTTTATTATGTAACAAGAGATTTGGCATGACATGTTCTGTATG 798

QY 2675 TTTTCAGGGAAAAATGCTCTTTAATGCTTTTTCAGGAACCTAACACAGTTATTTCCTATACCTGG 2734
|||||
Db 799 TTTTCAGGGAAAAATGCTCTTTAATGCTTTTTCAGGAACCTAACACAGTTATTTCCTATACCTGG 858

Qy 2735 ATTTTAGGTCCTCAAGAACTGCTGCTGTTTAGGAATAGAATGTGCATGAAGCCTAAA 2794
 |||||
Dd 859 ATTTTAGGTCCTCAAGAACTGCTGCTGTTTAGGAATAGAATGTGCATGAAGCCTAAA 918

QY 2795 TACCAAGAAAGCTTATCTACTGAATTTAAGCAAAGAAATAAAGGAGAAAGAGAGAATCTG 2854

Db 919 TACCAAGAAAGCTTATCTACTGAATTTAAGCAAAGAAATAAAGGAGAAAGAGAGAATCTG 978

QY 2855 AGAATTGGGAGGCATAGATTCTTATAAAAATCACAAAATTTGTTGTAATTTAGAGGGCA 2914
|||||
Db 979 AGAATTGGGAGGCATAGATTCTTATAAAAATCACAAAATTTGTTGTAATTTAGAGGGCA 1038

QY	2915	2914
GAAATTTAGAAATTAGTATAAAAAGCAGAAATTAGTATAGACTACATTCATTAAACATTT	2914	
Db	1039	1098
GAAATTTAGAAATTAGTATAAAAAGCAGAAATTAGTATAGACTACATTCATTAAACATTT	1039	1098

QY 2975 TTGTCAGGATATTTCCTCCGTAAACGGTAGTGAG--CACTCCTCATATACTAANT--ACTG 3030
|||||
Db 1099 TTGTCAGGATATTTCCTCCGTAAACGGTAGTGAGCGCACTTTTCATATACTAANTTATTGTTG 1158

QY 3031 TACATTAACTTGTATATAACAGAAATCTAAATATATATTTAATGAATTCAGCAATATA- 3089

Db 1159 TACATTAACTTGTATATAACAGAAATCTAAATATATATTTAATGAATTCAGCAATATAT 1218

	CAC	TGACCAGAAATTTGGAAATTTCCTCGCCGGGTATATATACCAGATGAGTAC	3149	
QY	3090	CAC	TGACCAGAAATTTGGAAATTTCAAAAATGTTTGTCGCCGGGTATATATACCAGATGAGTAC	3177
Dd	1219	CAC	TGACCAGAAATTTGGAAATTTCAAAAATGTTTGTCGCCGGG-TATATACCAGATGAGTAC	1277

QY 3150 AGTGAAGTAG-TTTATGTATCACCACACTGGGTTATTGCCAAGTTATATATCATCAAAAGC 3208
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1278 AGTGAGTAGTTTATGTATCACCACACTGGGTTATTGCCAAGTTATATATCATCAAAAGC 1337

Qy 3209 TGATGACTGGATCTTCTGTTACCTGGTTACAAATTATCAGAGTAGTAAACTTTGA 3268
=====

Db 1338 TGATGACTGGATGTTCTGTTACCTGGTTACAAATTATCAGAGTAGTAAACTTTGA 1397

QY 3269 TATATATCAGGATATTAAACTACACTAAGTATCATTTGATTCGATTCAGAAGTACTTTT 3328

Db 1398 TATATATCAGGATATTAAACTACACTAAGTATCATTTGATTCGATTCAGAAGTACTTTT 1457

QY 3329 GATATCTCTCAGTGCCTTCAGTGCATCATTTGTGAGCAATTGTC-TTTATATACGGTACTG 3387
=====

Db 1458 GATATCTCTCAGTGCCTTCAGTGCATCATTTGTGAGCAATTGTC-TTTATATACGGTACTG 1517

Accession	Sequence	Length
QY 3388	TAGCCATCTAGGCTGCTGTGGCATTCTCTAGATGTTCTTTTACACATAAATTC	3447
Db 1518	TAGCCATCTAGGCTGCTGTGGCATTCTCTAGATGTTCTTTTACACATAAATTC	1577

UY 3448 CTTATATCAGCTTG 3461
|||||
Db 1578 CTTATATCAGCTTG 1591

RESULT 9
AAT99070
ID AAT99070 standard; cDNA; 1310 BP.

KW Oestrogen regulated gene; pLIV1 gene partial sequence; breast cancer;

KW metastatic risk assessment; hormone therapy responsiveness;
KW tumour oestrogen receptor expression; lymph node involvement; ds.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1296
FT /*tag= a
XX
XX US5693465-A.
XX
XX 02-DEC-1997.
XX
XX 22-SEP-1994; 94US-0311023.
XX
XX 22-SEP-1994; 94US-0311023.
XX
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX
XX Gee JMW, Green CD, Manning DL, Nicholson RI;
XX
XX WPI; 1998-031740/03.
DR P-PSDB; AAW34528.
XX
XX
XX Assessment of metastatic risk or oestrogen responsive-ness in breast
XX cancer - by detecting expression of p191 protein
XX
XX Claim 20; Column 5-10; 11pp; English.
XX
XX This sequence represents a partial sequence of the p191 gene. The
XX protein encoded by this sequence is an oestrogen regulated gene
XX associated with breast cancer. The protein encoded by this sequence can
XX be used in the method of the invention. The method is for determining the
XX risk of metastasis of a female breast tumour and predicting the
XX responsiveness of a female breast tumour to hormone treatment
XX comprising if a tissue sample from the tumour expresses a polypeptide
XX determining at least 14 continuous amino acids of the protein encoded by
XX this sequence. The method is useful for determining the risk of
XX metastasis of a female breast tumour and to predict the responsive of the
XX female breast tumour to hormone treatment. p191 expression is highly
XX predictive of tumour oestrogen receptor expression and lymph node
XX involvement.
XX
XX Sequence 1310 BP; 354 A; 271 C; 287 G; 398 T; 0 other;
XX
XX
XX Query Match 37.2%; Score 1288; DB 19; Length 1310;
XX Best Local Similarity 99.8%; Pred. No. 1.6e-310;
XX Matches 1310; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
XX
XX 1119 GCTTTTATAGCCATTCCCATCATCATAGTTTCTCTGCTGGGGTTATCTTAGTGCCT 1178
XX
XX 1 GGGTTTATAGCCATTTCATCATCATAGTTTCTCTGCTGGGGTTATCTTAGTGCCT 60
XX
XX 1179 CTCATCAATCGGGTCTTTTTCAAATTTCTCTGCTGGGGTTATCTTAGTGCCT 1178
XX
XX 61 CTCATGAATCGGGTCTTTTTCAAATTTCTCTGCTGGGGTTATCTTAGTGCCT 1238
XX
XX 1239 ACTTTGAGTGGTATGCTTTTTCACCTTTCTCCACATCTCTGCTGGGGTTATCTTAGTGCCT 1298
XX
XX 121 ACTTTGAGTGGTATGCTTTTTCACCTTTCTCCACATCTCTGCTGGGGTTATCTTAGTGCCT 180
XX
XX 1299 ACTCATAGCCATGAAGAACAGCAATGAAATGAAAGAGGACCACTTTTCAGTCACTG 1358
XX
XX 181 AGTCATAGCCATGAAGAACAGCAATGAAATGAAAGAGGACCACTTTTCAGTCACTG 240
XX
XX 1359 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTTCACCTTTCTCCACATCTCTGCTGGGGTTATCTTAGTGCCT 1418
XX
XX 241 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTTCACCTTTCTCCACATCTCTGCTGGGGTTATCTTAGTGCCT 300
XX
XX 1419 CTAGAGGCTGATTTTCATGTTTCTGTTGAACATGCTCCTCAGATTTGATCAACAAATTT 1478
XX
XX 301 CTAGAGGCTGATTTTCATGTTTCTGTTGAACATGCTCCTCAGATTTGATCAACAAATTT 360

QY 1479 AAAGATAAGAGAAAGAAAGAAATCAGAGAAACCTGAAATGATGATGATGCGAGATTAAAG 1538
DB
361 AAAGATAAGAGAAAGAAAGAAATCAGAGAAACCTGAAATGATGATGATGCGAGATTAAAG 420
QY 1539 AAGCAGTTGTCCTCAAGTATGAATCTCAACTTTCAACAATGAGGAGAAAGTAGATACAGAT 1598
DB
421 AAGCAGTTGTCCTCAAGTATGAATCTCAACTTTCAACAATGAGGAGAAAGTAGATACAGAT 480
QY 1599 GATCGAACTGAAGGCTATTTACGAGCAGACTCACAAGGCCCTCCCACTTTGATTTCTCAG 1658
DB
481 GATCGAACTGAAGGCTATTTACGAGCAGACTCACAAGGCCCTCCCACTTTGATTTCTCAG 540
QY 1659 CAGCCTGCAGTCTTGGAGAGAGAGGCTCATGATAGTCTCATCTCCACAGGAAGTTC 1718
DB
541 CAGCCTGCAGTCTTGGAGAGAGAGGCTCATGATAGTCTCATCTCCACAGGAAGTTC 600
QY 1719 TACAATGAATATGTACCCAGAGAGGTCGAAGAATAAATGCCATTCACATTTCCACAGTACA 1778
DB
601 TACAATGAATATGTACCCAGAGAGGTCGAAGAATAAATGCCATTCACATTTCCACAGTACA 660
QY 1779 CTCGCCAGTCAGACGATCTCATTTCCACCATCATGACTTACCATCATATTTCTCCATCAT 1838
DB
661 CTCGCCAGTCAGACGATCTCATTTCCACCATCATGACTTACCATCATATTTCTCCATCAT 720
QY 1839 CACCACCAACCAACCAACCACTCTCAGAGCAGCGCTACTCTCGGAGGAGGCTG 1898
DB
721 CACCACCAACCAACCAACCACTCTCAGAGCAGCGCTACTCTCGGAGGAGGCTG 780
QY 1899 AAAGATGCGGGCTCGCCACTTTGGCTGGATGGTGGTGAATGGGCTGAGCGCTGCACAAAT 1958
DB
781 AAAGATGCGGGCTCGCCACTTTGGCTGGATGGTGGTGAATGGGCTGAGCGCTGCACAAAT 840
QY 1959 TTCAGGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTGGTTAAAGT 2018
DB
841 TTCAGGATGGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTGGTTAAAGT 900
QY 2019 ACTTCTGTTGCTGTTGTTCTGTCATGAGTGGCTCATGAATAGGAGTGGTGGTGGTCTA 2078
DB
901 ACTTCTGTTGCTGTTGTTCTGTCATGAGTGGCTCATGAATAGGAGTGGTGGTGGTCTA 960
QY 2079 CTAAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTAATGATGATGTCAGCCATGCTG 2138
DB
961 CTAAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTAATGATGATGTCAGCCATGCTG 1020
QY 2139 GCGTATCTTGAATGGCAACAGGAATTTTCATTTGCTCATATTGCTGAAATGTTTCTATG 2198
DB
1021 GCGTATCTTGAATGGCAACAGGAATTTTCATTTGCTCATATTGCTGAAATGTTTCTATG 1080
QY 2199 TGGATATTTGCACTTACTGCTGGCTTATTCATGATGATGCTGCTGCTGCTGCTGCTGCT 2258
DB
1081 TGGATATTTGCACTTACTGCTGGCTTATTCATGATGATGATGCTGCTGCTGCTGCTGCTGCT 1140
QY 2259 GAAATGCTGCAATGATGCTGAGTACCATGATGAGCGCTGGGGTATTTCTTTTAA 2318
DB
1141 GAAATGCTGCAATGATGCTGAGTACCATGATGAGCGCTGGGGTATTTCTTTTAA 1200
QY 2319 CAGATGCTGGGATGCTTTTGGGTTTGGAAATGATGTTACTTATTTCCATTTTGAACAT 2378
DB
1201 CAGATGCTGGGATGCTTTTGGGTTTGGAAATGATGTTACTTATTTCCATTTTGAACAT 1259
QY 2379 AAAATCGTGTTCGTATAAAATTTCTAGTTAAGGTTTAAATGCTAGAGTAGCT 2430
DB
1260 AAAATCGTGTTCGTATAAAATTTCTAGTTAAGGTTTAAATGCTAGAGTAGCT 1310

RESULT 10
AAT33219
ID AAT33219 standard; cDNA; 1310 BP.
XX
AC AAT33219;
XX
DT 16-OCT-1996 (first entry)
XX

Tue Nov 26 12:43:33 2002

Oestrogen-regulated pLIV1 gene partial cDNA clone.
Oestrogen-regulated gene; pLIV1; breast cancer; tumour; diagnosis;
marker; metastasis; ss.
Homo sapiens.
Key Location/Qualifiers
CDS 1..1296
/*tag= a
XX CA2132500-A.
XX 21-MAR-1996.
XX 20-SEP-1994; 94CA-2132500.
XX 20-SEP-1994; 94CA-2132500.
XX (UYMA-) UNIV WALES COLLEGE OF MEDICINE.
XX Gee JM, Green CD, Manning DL, Nicholson RI;
XX WPI: 1996-268987/28.
XX P-PSDB; AAR98004.
XX Oestrogen-regulated pLIV1 gene and deduced polypeptide - useful for
XX predicting the propensity for metastatic spread or the
XX responsiveness to endocrine treatment of breast tumour
XX Claim 20; Page 12-14; 27pp; English.
XX The oestrogen-regulated pLIV1 gene (partial cDNA given in AAT33219)
XX is a candidate for the hormonal regulation of tumour invasion of
XX breast cancer, a prognostic marker of metastatic spread, and a
XX suitable target for therapeutic intervention by anti-hormones and
XX antibody-directed methods. It is significantly associated with
XX oestrogen-dependent (ER+) disease. The pLIV1 coding sequence was
XX isolated from a series of oestrogen-regulated genes. A 3' extended
XX version (AAT33220) of the partial cDNA clone was also dectd. The
XX polypeptide (AAR98004) encoded by the partial cDNA contains 3
XX immunogenic regions.
XX Sequence 1310 BP; 354 A; 273 C; 285 G; 398 T; 0 other;
XX
XX Query Match 37.1%; Score 1284.8; DB 17; Length 1310;
XX Best Local Similarity 99.7%; Pred. No. 1e-309;
XX Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
XX
XX 1119 GGTGTTATAGCCATTTCCATCATCAGTTTCCCTCTCTCTGCTGGGGTATATCTTAGTGCCT 1178
XX 1 GGTGTTATAGCCATTTCCATCATCAGTTTCCCTCTCTCTGCTGGGGTATATCTTAGTGCCT 60
XX
XX 1179 CTCATGAATCGGGTGTGTTTCAAAATTTCTCCTCAGTTTCTTGGGACTGGCCGTGGG 1238
XX 61 CTCATGAATCGGGTGTGTTTCAAAATTTCTCCTCAGTTTCTTGGGACTGGCCGTGGG 120
XX
XX 1239 ACTTGTAGTGTGATGCTTTTACACCTTCTTCCACATTTCTCATGCAAGTCAACCCAT 1298
XX 121 ACTTGTAGTGTGATGCTTTTACACCTTCTTCCACATTTCTCATGCAAGTCAACCCAT 180
XX
XX 1299 ACTCATAGCCATGAAGAACACAGCAATGGAATGAAGAGAGACACCTTTTCAGTCATCTG 1358
XX 181 AGTCATAGCCATGAAGAACACAGCAATGGAATGAAGAGAGACACCTTTTCAGTCATCTG 240
XX
XX 1359 TCTTCTCAAAACATAGAAGAAAGTGCCTATTTGATTTCCAGTGAAGGGTCTAACAGCT 1418
XX 241 TCTTCTCAAAACATAGAAGAAAGTGCCTATTTGATTTCCAGTGAAGGGTCTAACAGCT 300
XX
XX 1419 CTAGAGGGCTCTATTTCATGTTTCTTGTGAACATGCTCCTCACATTTGATCAACAAATTT 1478
XX 301 CTAGAGGGCTCTATTTCATGTTTCTTGTGAACATGCTCCTCACATTTGATCAACAAATTT 360

1479 AAAGATAAGAGAAAAAGAAATCAGAGAAACCTGAAATCATGATGATGAGATTAAAG 1538
1480 AA
361 AAAGATAAGAGAAAAAGAAATCAGAGAAACCTGAAATCATGATGATGAGATTAAAG 420
1539 AAGCAGTGTCCCAAGTATGAATCTCAACTTTCAACAAATGAGGAGAAAGTAGATACAGAT 1598
1540 AA
421 AAGCAGTGTCCCAAGTATGAATCTCAACTTTCAACAAATGAGGAGAAAGTAGATACAGAT 480
1599 GATCGAACTGAAGGCTATTTACGAGAGAGCTCACAAGAGCCCTCCCACTTTGATTCTCAG 1658
1600 AA
481 GATCGAACTGAAGGCTATTTACGAGAGAGCTCACAAGAGCCCTCCCACTTTGATTCTCAG 540
1659 CAGCCCTCAGTCTTGGAAAGAGAGAGGTCATGATAGTCATGCTCATGCCAGAGAACTC 1718
1660 AA
541 CAGCCCTCAGTCTTGGAAAGAGAGAGGTCATGATAGTCATGCCAGAGAACTC 600
1719 TACAATGATATGTATACCCAGAGGGTGCAGAGAAATAAATGCCATTCCACCATACA 1778
1720 AA
601 TACAATGATATGTATACCCAGAGGGTGCAGAGAAATAAATGCCATTCCACCATACA 660
1779 CTCGGCCAGTCAGAGATCTCATTCACCAACCATCATGACTACCATCATATTTCTCCATCAT 1838
1780 AA
661 CTCGGCCAGTCAGAGATCTCATTCACCAACCATCATGACTACCATCATATTTCTCCATCAT 720
1839 CACCACCAACCAACCAACCATCTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGCTG 1898
1840 AA
721 CACCACCAACCAACCAACCATCTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGCTG 780
1899 AAAGATGCGGGCTCGCCACTTTGGCTGGATGTTGATTAATGGTGTGATGGCTTGCACAAAT 1958
1900 AA
781 AAAGATGCGGGCTCGCCACTTTGGCTGGATGTTGATTAATGGTGTGATGGCTTGCACAAAT 840
1959 TTACAGCATGGCTTAGCAATTTGGTGTCTTTTACTGAAGGCTTATCAAGTGGTTAACT 2018
1960 AA
841 TTACAGCATGGCTTAGCAATTTGGTGTCTTTTACTGAAGGCTTATCAAGTGGTTAACT 900
2019 ACTTCTGTGCTGTGTTCTGTCTGATGCTTGCCTCATCAATTAGTGTGCTTGTGTTCTA 2078
2020 AA
901 ACTTCTGTGCTGTGTTCTGTCTGATGCTTGCCTCATGATGCTTGTGCTGTTCTA 960
2079 CTAAGGCTGCGATGACCGTTTAAGCAGGCTGCTCTTATTAATGATGCTGACGCTATGCTG 2138
2080 AA
961 CTAAGGCTGCGATGACCGTTTAAGCAGGCTGCTCTTATTAATGATGCTGACGCTATGCTG 1020
2139 GCGTATCTTGAATGGGCAACAGGAATTTTCATTTGCTCATTTATGCTGAAATGTTTCTATG 2198
2140 AA
1021 GCGTATCTTGAATGGGCAACAGGAATTTTCATTTGCTCATTTATGCTGAAATGTTTCTATG 1080
2199 TGGATATTGCACTTACTGCTGGCTTATTCATGATGTTGCTGCTGGTGTGATGCTACCT 2258
2200 AA
1081 TGGATATTGCACTTACTGCTGGCTTATTCATGATGTTGCTGCTGGTGTGATGCTACCT 1140
2259 GAAATGCTGCAATGATGCTGACCATGATGACCATGATGACCATGATGACCATGATGACCAT 2318
2260 AA
1141 GAAATGCTGCAATGATGCTGACCATGATGACCATGATGACCATGATGACCATGATGACCAT 1200
2319 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATGATGTTACTTATTATTCATATTTGAAAT 2378
2320 AA
1201 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATGATGTTACTTATTATTCATATTTGAAAT 1259
2379 AAAATGCTGTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTAGCT 2430
2380 AA
1260 AAAATGCTG-TTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTAGCT 1310
2430 AA
RESULT 11
AAF16156
ID AAF16156 standard; cDNA; 1193 BP.
XX
XX AAF16156;
XX
XX 13-MAR-2001 (first entry)
XX

Tue Nov 26 12:43:33 2002

us-09-642-034-l.l.rng

KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0218290.
XX 14-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.


```

F11 /transl_except= (pos:879..881,aa:Xaa)
F11 /note= "Xaa = unknown; the CDS does not include a
F11 stop codon"

```

```

F1 535..537
Ft /*tag= b
Ft /note= "encodes Lys"
Ft 538..540
Ft /*tag= c
Ft /note= "encodes Lys"
Ft

```

XX PN WO200216939-A2.

XX PD 28-FEB-2002.

XX PF 20-AUG-2001: 2001WO-US25997.

XX 18-AUG-2000: 2000US-0642034.

PR 08-DEC-2000; 2000US-0733320.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Mack D, Gish KC, Wilson KE;

DR WPI; 2002-242033/29.

DR P-PSDB; AAM31199.
XX

PT Screening for drugs that affect expression of the breast cancer 4 gene
PT or its fragments, use of these to treat prostate and breast cancer, and
PT diagnosing these diseases -

PS Disclosure; Fig 3; 83pp; English.

The presence sequence is that of a partial cDNA for the murine breast cancer 4 gene (BCR4). The cDNA encodes the protein given in AAM51199. The invention also provides human BCR4 cDNA (see AAM9299) and protein (see AAM51198) sequences. Methods are claimed for using BCR4 polynucleotides, polypeptides and antibodies to: screen drug candidates; screen bioactive agents capable of binding to BCR4; evaluate the effect of a prostate cancer and/or breast cancer drug; diagnose prostate cancer or breast cancer; screen for a bioactive agent capable of interfering with the binding of BCR4 to an antibody which binds BCR4; neutralise the effect of BCR4; treat prostate cancer or breast cancer using an inhibitor of BCR4 (e.g. an antibody); localise a therapeutic moiety to, or treat, prostate cancer or breast cancer tissue by conjugating an antibody for BCR4 to a cytotoxic agent or radioisotope; inhibit prostate cancer or breast cancer by administering an antisense molecule to the present nucleic acid sequence; elicit an immune response by administering BCR4 or a nucleic acid encoding BCR4; and determine the prognosis of an individual with prostate cancer or breast cancer by determining the level of BCR4 in a sample, a high level indicating a poor prognosis. A biochip comprising BCR4 nucleic acids is also claimed.

XX
Sequence 899 BP: 252 A: 239 C: 198 G: 203 T: 7 other;

Query Match	17.9%	Score 619.6;	DB 24;	Length 899;
Best Local Similarity	84.3%;	pred. NO. 4.8e-144;		
Best Local Conservative	1.752;	Wismarches 131;		
		Indels	8;	Gaps 5;

Qy 960 GGCATGGGCATCCAGGTTCCGCTGAATGCACAGAGTTCAACTATCTCTGTCCAGCCATC 1019

Qy 1020 ATCAACCAATTGATGCTAGATCTTGTCTGATTCTATA--CAAGTGAAAGAAGGCTGAA 1076

QY 1077 ATCCCTCCAAGACCTATTTCATTACAATAAGCTGGTGGTTGATGCCATTTC 1136

QY 1137 ATCATCAGTTTCC^TGTCTCT^CTCGCTGGGGGTTAT^TCTTAGTGCCCTC^ATGAATCGGGRGTTT 1196

Db	181	 ATCATCAGTTTCTGTCTCTGTCTGGAGTTCATCTTGTGTGCCACTCATGAACCGGTATTTT	240
Qy	1197	TTCAAAATTTCTCTGAGTTTCTTGTGCACTGGCCGTTGGAGCTTTTGAGTGGTATGCT	1256
Db	241	TTCAAGTTCTCTGCTGAGCTTCTCTGTCGCGCTGGCCGTCGGAACGCTGAGTGGCGATGCT	300
Qy	1257	TTTTTACACCTTCTTCCACATTTCTCATGCAAGTCAACCACTAGTCATAGCCATGAAGAA	1316
Db	301	CTGTTACATCTTCTCCACACTCTCATCAAGTCTCATGACACAGTCATAGCCCATGAAGAG	360
Qy	1317	CCAGCAATGGAATGAAAAGAGGACACTTTTTCAGTCTATCTGTCTTCTCAAAACATAGAA	1376
Db	361	CCAGCGATGGAATGAAAAGAGCCCTCTGTTCAACCCACTCTGCGCTCAGAATATAGAA	420
Qy	1377	GAAAGTGCCTATTTTGAATTCACAGTGGGAAGGCTTAACAGCTCTAGAGGCCCTGTATTTC	1436
Db	421	GAAAGTCCCTATTTTGAATTTCCAGTGGGAAGTCTGACGGCTCTAGGGGCTTATATTTC	480
Qy	1437	ATGTTTCTTCTTCAACATGTCTTCACATGTGATCAACAATTTAAAGATAAGAAGAAAAAG	1496
Db	481	ATGTTTCTTGTGGACACGTACTCACATGTATCAAGCAATTTAAAGATAGAAGAAANAN	540
Qy	1497	AATCAGAAGAAACCTGAAAATGATGATGATGATGGAATTAAGAAGCAGTTGTCCAAGTAT	1556
Db	541	AANMAAAGAAACCTGAAAATGATGAGGATGTGGAGAGCAAGACGAGCTGTCCAATATC	600
Qy	1557	GAATCTCACTTTCAACAAATGAGGAGAAGTAGATACAGATGATCGAAGCTGAAGGCTAT	1616
Db	601	GACTCTCAGTCTTCTCTCAATGAAGAAGAGTGGACCCAGGGGAACGACCTGAAAGCTAT	660
Qy	1617	TTACGAGCAGACTCACAAAGAGCCCTCCCACTTTTGATTTCTCAGCAGCCGTG-CAGTCTTGGGA	1675
Db	661	CTGCGAGCCGACTCCCAAGAGCCCTCCCTCTTGATTTCCACAGCAGCCGACGATGTGGGA	720
Qy	1676	AGAGAGAGAGGTGATGATAGTCTATGCTCA-TCCACAGGAAGTCTACAAATCAATATG-T	1732
Db	721	AGAGGAAGANGTCTATGATGCCCATGCACACCCCAAGAAGTCTACAAATGAATATGGTG	780
Qy	1733	ACCCAGAGGGTGCAGAATAAATGCCATTTCACATTTCCACGATACACTCGGCCAGTCAGA	1792
Db	781	CCCAGGGGGTGCAGAACAAGTGCCATTTCACACTTTCACGATACGCTGGCCAGTCGGA	840
Qy	1793	CGATCTCATTCACCACCATCATGACTACCATCATATTTCCATCATCATCACCAC	1844
Db	841	CGACCT-ATNCACCACCATTACGACTACCATCATCATTTCTGNACCACCACCTC	891
RESULT 14			
AADI3487			
ID	AADI3487 standard; cDNA; 551 BP.		
XX	XX		
AC	AADI3487;		
XX	XX		
DT	06-NOV-2001 (first entry)		
XX	XX		
DE	Human LIV-1-164647 cDNA fragment (nucleotides 1690-2240).		
XX	XX		
KW	Human; LIV-1-164647; cytosolic; estrogen-inducible gene; tumour;		
KW	cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;		
KW	salivary gland; carcinoma; drug screening; therapy; ss.		
OS	Homo sapiens.		
XX	XX		
PN	W0200155178-A2.		
XX	XX		
PD	02-AUG-2001.		
XX	XX		
PF	25-JAN-2001; 2001WO-US02622.		
XX	XX		
PR	25-JAN-2000; 2000US-0177951.		
PR	10-APR-2000; 2000US-0195761.		
XX	XX		

Seq. search results -
file copy

DB 1021 TCAACG
943 GCATG
961 GC
883

Db 3221 CAAAATTATCAGAGTAGTAAACCTTTTGATATATATAGGAGTATTTAAACCTACACATAGTA 3280

QY 3301 TCATTTGATTTCGATTCAGAAAAGTACTTTTGATATCTCTCAGTGTCTCAGTGTCTATCATTTCT 3360

Db 3281 TCATTTGATTTCGATTCAGAAAAGTACTTTTGATATCTCTCAGTGTCTCAGTGTCTATCATTTCT 3340

QY 3361 GAGCAATTCCTTTTATATACGGTACTGTAGCCATACATAGCCCTGTCTGTGGCATTTCTCTA 3420

Db 3341 GAGCAATTCCTTTTATATACGGTACTGTAGCCATACATAGCCCTGTCTGTGGCATTTCTCTA 3400

QY 3421 GAGTTTCTTTTATACCAATAAAATTCCTTATATCAGCTTG 3461

Db 3401 GAGTTTCTTTTATACCAATAAAATTCCTTATATCAGCTTG 3441

RESULT 3
AX207207 2776 bp DNA linear PAT 30-AUG-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
Goddard A, Gurney A L., Smith V., Hongo J A. and de Sauvage F.
Compositions and methods for treatment of cancer
Patent: WO 0155178-A 3 02-AUG-2001;
GENENTECH, INC. (US)
FEATURES
Location/Qualifiers
1..2776
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 823 a 583 c 592 g 778 t
ORIGIN
Query Match 79.8%; Score 2761.2; DB 6; Length 2776;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2766; Conservative 0; Mismatches 8; Indels 0; Gaps 0
QY 68 GCCCTGTGGAAACCAACCTCGCGCGTGGCGCGTGGGAAACAGAGCGCGCGGAGA 127
Db 3 GCCCTGTGGAAACCAACCTCGCGCGTGGCGCGTGGGAAACAGAGCGCGCGGAGA 62
QY 128 CGAAGCGCAATGCGGAGGAGTATCTGTAATCTTGATCTGACCTTTGCCCTCTCTCT 187
Db 63 CGAAGCGCAATGCGGAGGAGTATCTGTAATCTTGATCTGACCTTTGCCCTCTCTCT 122
QY 188 CACAATCCCCCTTCATGAACATAAAGAGAGCTGCTTTCCCGCAGACCACTCAGAAAAATTAG 247
Db 123 CACAATCCCCCTTCATGAACATAAAGAGAGCTGCTTTCCCGCAGACCACTCAGAAAAATTAG 182
QY 248 TCCGAATTTGGGAATCTGGCAATTAATGTTGAGTTTGGCAATTTCCACAGCGCAATCATCT 307
Db 183 TCCGAATTTGGGAATCTGGCAATTAATGTTGAGTTTGGCAATTTCCACAGCGCAATCATCT 242
QY 308 ACACAGCTTTCTACCGCTATGGAGAAAATAATTTCTTTGTGACCTTCAAGGGTTTCAGAAA 367
Db 243 ACACAGCTTTCTACCGCTATGGAGAAAATAATTTCTTTGTGACCTTCAAGGGTTTCAGAAA 302
QY 368 ATTACTTCAAAATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATGACCATCA 427
Db 303 ATTACTTCAAAATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATGACCATCA 362
QY 428 CCATCTCAGACGAGCAGCATCTCAGACCATGAGCTGCTCAGACCATGAGCATCA 487
Db 363 CCATCTCAGACGAGCAGCATCTCAGACCATGAGCTGCTCAGACCATGAGCATCA 422
QY 488 CTCAGACGAGCAGCATCTCAGACCATGAGCTGCTCAGACCATGAGCATCA 547
Db 423 CTCAGACGAGCAGCATCTCAGACCATGAGCTGCTCAGACCATGAGCATCA 482

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 8, 2002, 16:42:31 ; Search time 6148.03 Seconds
(without alignments)
16383.279 Million cell updates/sec

Title: US-09-642-034-1
Perfect score: 3461
Sequence: 1 ctcgtgccgaattccgacg.....aaattcttatatcagcttg 3461

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pt.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3461	100.0	3461	6	AX465588	AX465588 Sequence
2	3391	98.0	3461	6	AX207205	AX207205 Sequence
3	2761.2	79.8	2776	6	AX207207	AX207207 Sequence
4	2716	78.5	2744	9	HSU41060	U41060 Homo sapien
5	2249	65.0	2404	6	I76892	I76892 Sequence 3
6	2134.2	61.7	2236	9	BC008317	BC008317 Homo sapi
7	1753.2	50.7	3287	10	AB071697	AB071697 Homo sapi
8	1486.8	43.0	1597	6	AX017261	AX017261 Sequence
9	1288	37.2	1310	6	I76891	I76891 Sequence 1
10	1137.8	32.9	221941	9	AC091060	AC091060 Homo sapi
11	1137.8	32.9	224788	2	AF001905	AF001905 Homo sapi
12	804	23.2	160170	2	AF001158	AF001158 Homo sapi
13	619.6	17.9	899	6	AX465590	AX465590 Sequence
14	549.4	15.9	551	6	AX207216	AX207216 Sequence
15	429.2	12.4	193437	2	AC126667	AC126667 Rattus no
16	390.8	11.3	395	6	AX158237	AX158237 Sequence
17	336.8	9.7	445	11	G24009	G24009 human STS W
18	319.2	9.2	337	6	AX393861	AX393861 Sequence
19	319	9.2	5231	9	AB033091	AB033091 Homo sapi
20	319	9.2	5322	6	AX405756	AX405756 Sequence
21	282.8	8.2	61103	2	AC131276	AC131276 Homo sapi
22	253.8	7.3	391	6	AX207209	AX207209 Sequence
23	240	6.9	241	6	AX303040	AX303040 Sequence
24	191.2	5.5	193437	2	AC126667	AC126667 Rattus no
25	165	4.8	177	6	AX106381	AX106381 Sequence
26	165	4.8	177	6	AX140672	AX140672 Sequence
27	165	4.8	177	6	AX200532	AX200532 Sequence
28	165	4.8	177	6	AX267188	AX267188 Sequence
29	144	4.2	52684	2	AC014796	AC014796 Drosophil
30	144	4.2	174712	3	AC010041	AC010041 Drosophil
31	144	4.2	195148	3	AC010019	AC010019 Drosophil
32	144	4.2	282267	3	AE003555	AE003555 Drosophil
33	129.2	3.7	2174	10	BC021530	BC021530 Mus muscu
34	128.4	3.7	2015	9	AB051127	AB051127 Macaca fa
35	127.6	3.7	4573	9	AX409002	AX409002 Sequence
36	127.6	3.7	4573	9	HUMORFKG1P	D31887 Human mRNa
37	126	3.6	4610	6	AX078294	AX078294 Sequence
38	126	3.6	4698	6	AX281784	AX281784 Sequence
39	123.4	3.6	1383	6	AX118897	AX118897 Sequence
40	123.4	3.6	1803	9	AF193052	AF193052 Homo sapi
41	123.4	3.6	2604	9	AK027652	AK027652 Homo sapi
42	123.4	3.6	3169	9	BC012125	BC012125 Homo sapi
43	123.4	3.6	3246	9	AB040120	AB040120 Homo sapi
44	122.6	3.5	1895	9	BC027884	BC027884 Homo sapi
45	121.4	3.5	55577	2	AC014109	AC014109 Drosophil

ALIGNMENTS

RESULT 1
AX465588
LOCUS AX465588 3461 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0216939.
ACCESSION AX465588
VERSION AX465588.1 GI:21899891
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mack,D., Gish,K.C. and Wilson,K.E.
TITLE Methods of diagnosis of cancer and screening for cancer modulators
JOURNAL Patent: WO 0216939-A 1 28-FEB-2002;